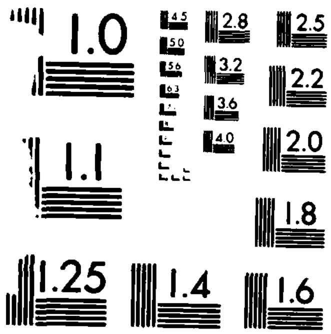
AD-A165 038 SPACE HANDBOOK(U) AIR MINISTRY LONDON (ENGLAND) 1/4
C D COCHRAN ET AL. JAN 85 AU-18

UNCLASSIFIED F/G 15/7 NL

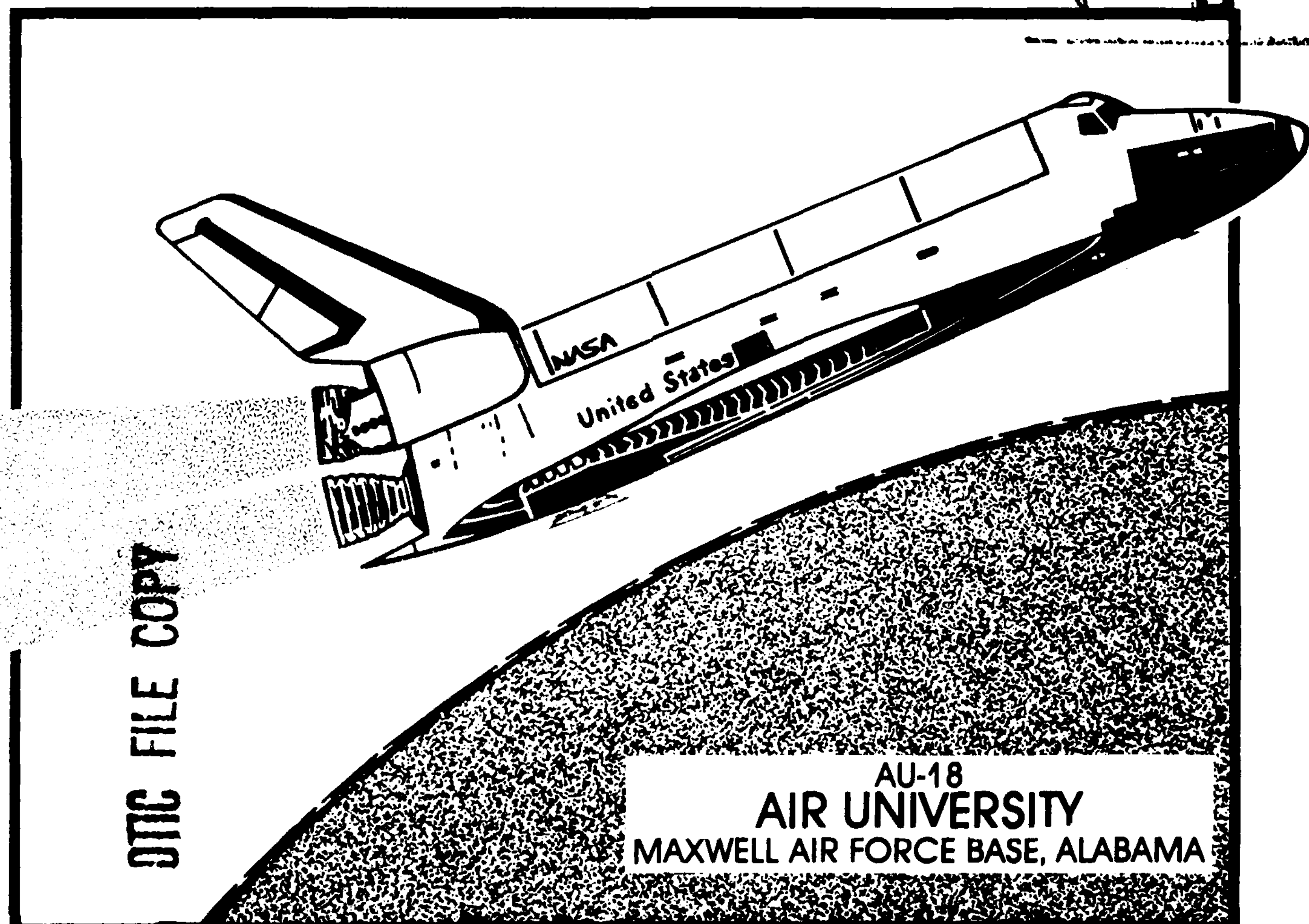


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A165 038

SPACE

HANDBOOK



AU-18
AIR UNIVERSITY
MAXWELL AIR FORCE BASE, ALABAMA

AU-18

Space Handbook

Prepared by
AIR COMMAND AND STAFF COLLEGE

Edited by
Lt Col Curtis D. Cochran
Lt Col Dennis M. Gorman
Maj Joseph D. Dumoulin

AIR UNIVERSITY PRESS
AIR UNIVERSITY (AU)
MAXWELL AIR FORCE BASE, ALABAMA 36112-5532
January 1985

Twelfth Revision, January 1985

Eleventh Revision, August 1977
Tenth Revision, July 1974
Ninth Revision, July 1972
Eighth Revision, July 1970
Seventh Revision, July 1969
Sixth Revision, July 1968
Fifth Revision, July 1967
Fourth Revision, July 1966
Third Revision, July 1965
Second Revision, July 1964
First Revision, July 1963
First Edition, December 1962

This publication has been reviewed and approved by competent personnel of the preparing command in accordance with current directives on doctrine, policy, essentiality, propriety, and quality.

For sale by the Superintendent of Document
US Government Printing Office
Washington DC 20402

Contents

Chapter 3

Propulsion Systems

Chapter 4

Space Vehicle Electrical Power

Chapter 5

Guidance and Control

Chapter 6

Active US Space Launch Vehicles and Upper Stages

Chapter 7

Global Communications

Chapter 8

Atmospheric Penetration

Chapter 9

Directed Energy Systems

Chapter 10

Reliability of Space Systems

Chapter 11

Bioastronautics

Chapter 12

Military Space Systems

Chapter 13

Space Transportation System

Chapter 14

Space Support Organizations

Chapter 15

Space Policy and Doctrine

APPENDIXES

foreword

The Air Force has been in space for twenty-five years. In that short time space has evolved from an arena for technological experiments to the point that space systems are now an integral part of our military force structure. Our military forces depend on satellites for a wide range of critical functions—communications, weather forecasting, navigation, warning, and treaty verification. Further, an even greater role for military space activities is evident from our national space policy, the President's strategic defense initiative, and expanding operational requirements.

Historically, our professional military education system has been at the forefront in addressing new challenges. Space is no exception. Air University published the first *Space Handbook* in 1962, and has done a great job in keeping pace with the rapid technological changes characteristic of space systems and operations.

This *Handbook* is an excellent reference document. To the novice, it provides a basic understanding of the physical laws and principles of the space environment, and background into the evolution of space policy and doctrine. To the space professional, it provides a stimulus for new ideas and concepts.

JAMES V. HARTINGER
General, USAF
Commander

Preface

This text is prepared and revised* by the Electronic Warfare and Space Division of the Air Command and Staff College. The *Space Handbook* serves as a text for resident courses within Air University. As such, the text is written at an intermediate level of academic difficulty but with considerable depth of detail.

The text supports the following general objectives of the resident courses:

1. To provide the students with an introduction to the basic physical laws and principles which permit and limit space operations.

2. To provide the students familiarization with the objectives of the national space effort; current Soviet and US operations in space; and the organizations that support the US space effort.

3. To provide insights regarding the significance of space operations on military capability and to stimulate thought on new ideas and concepts so that the students may apply their knowledge in the performance of space planning and operational duties.

Recommendations for improvements of the *Space Handbook*, and request for copies, should be sent to ACSC/EDCW, Maxwell AFB AL 36112-5542.

*This twelfth revision is a partial update with major revisions to chapters 5, 6, 9, 13, and 15.

Chapter 1

THE SPACE ENVIRONMENT

For thousands of years people have looked at the heavens and wondered. What are the stars? What are the planets? The moon? The shooting stars? Why and how do these heavenly bodies move? Answers derived from superstition, philosophy, religion, and fear abound in the literature and folklore of all people. Only recently in the world's history have observation and experimentation provided answers. Even these answers are tentative. The success of the first manned lunar landing, Apollo 11, was truly a milestone in the search for understanding that continues at a quickening pace.

Today there are at least two ways of looking at the space environment. The first is the magnificent look—the look that sees space as the whole universe in terms of both matter and energy. The second is the practical look—the look that sees space as another region in which people have begun travel. The former staggers the imagination and stimulates both wonder and reverence. The latter is the immediate concern of the military.

This chapter will start with a brief discussion of the universe as we believe it to be today. We will present the near-earth space characteristics that have immediate military importance as well.

THE UNIVERSE

People live on an earth that is one of nine planets in separate orbits around a star called the sun. In addition to the planets, the Solar System contains more than 30 natural satellites, a variable number of artificial satellites, about thirty thousand asteroids, a hundred billion comets and countless specks of dust. These numbers seem impressive. But the sun, which is the master of the entire system, is more impressive. It contains 99.9 percent of the matter in the Solar System.

The nearest stellar neighbor to the sun is a star called Alpha Centauri, a conspicuous double star that is visible from the southern hemisphere. About two degrees from it is Proxima Centauri, so named because it was once thought to be even nearer than Alpha Centauri. This group of three stars is about 4.3 light years from the sun. That is, it takes light, traveling at 186,000 miles per second, about 4.3 years to reach the Solar System. It would take over 100,000 years for a spacecraft travelling at 25,000 miles per hour to make the trip. To date, the fastest people have ever travelled is about 25,000 miles per hour. It is clear that a person cannot live long enough to travel to the nearest star. This is the present status of such travel. In the future, greater speeds, or perhaps the contraction of time, which Dr Albert Einstein predicted, may make such travel a reality.

Another way to visualize these immense distances is to imagine the sun as represented by a golf ball. On this scale a pair of golf balls that are a fraction of a mile apart and 500 miles away would represent Alpha Centauri. Proxima Centauri would be a grain of sand about 25 miles from the pair of golf balls. Some of the other closest stars are:

Barnard's Star	6.1 light years
Wolf 359	8 light years
Sirius	8.6 light years

Procyon	11	light years
Vega	27	light years

But these are only the sun's nearest neighbors. Betelgeuse is 300 light years away. Polaris (the North Star) is 600 light years away. Yet all of these are only a few of a vast array of stars that form a group called the Milky Way.

The name given to a large group of stars, dust, and gas that stay together in a structure is a galaxy. The Milky Way is simply the view from earth of the galaxy in which the sun is one star. Figure 1-1 is a view of the Milky Way as it would appear from the side. Figure 1-2 is its appearance from above. These are artist's conceptions, because we cannot take such pictures from inside the galaxy. The Milky Way is approximately 100,000 light years in diameter. The Solar System is located in one of the spiral arms of the galaxy, about 30,000 light years from the center.

How many stars are there in this magnificent structure? No one knows the exact answer. Dr Harlow Shapley estimates the number at approximately 200 billion. All of these are in motion around the center of the galaxy. At the distance of the Solar System from the center of the galaxy, the speed of revolution is about 135 miles per *second* or about 486,000 miles per hour. Even at this tremendous speed, it takes the sun 220 million years to make one trip around the galactic center. Stars closer to the center move faster and those farther from the center move more slowly.

Beyond the Milky Way the telescopes show other objects. What are they? Gas? Dust? Stars? In the year 1755 Immanuel Kant suggested that these were "island universes"—other galaxies similar to the Milky Way, each consisting of billions of stars. It was not until 1917 that an astronomer using the Mount Wilson telescope identified a star in one of the objects beyond the Milky Way.

Today we know that these objects are other galaxies. Some have diameters in the order of 7,000 light years. Others have diameters even greater than the Milky Way—about 150,000 light years. The smaller ones probably contain a billion stars. The larger ones may contain two hundred billion or more. How many galaxies are there in the universe? The answer must be an estimate, and the most recent one is huge—one hundred billion! Finally, if one asks the question, "How many stars are there in the universe?" the answer is approximately 10^{21} stars! Although no one can imagine this number,

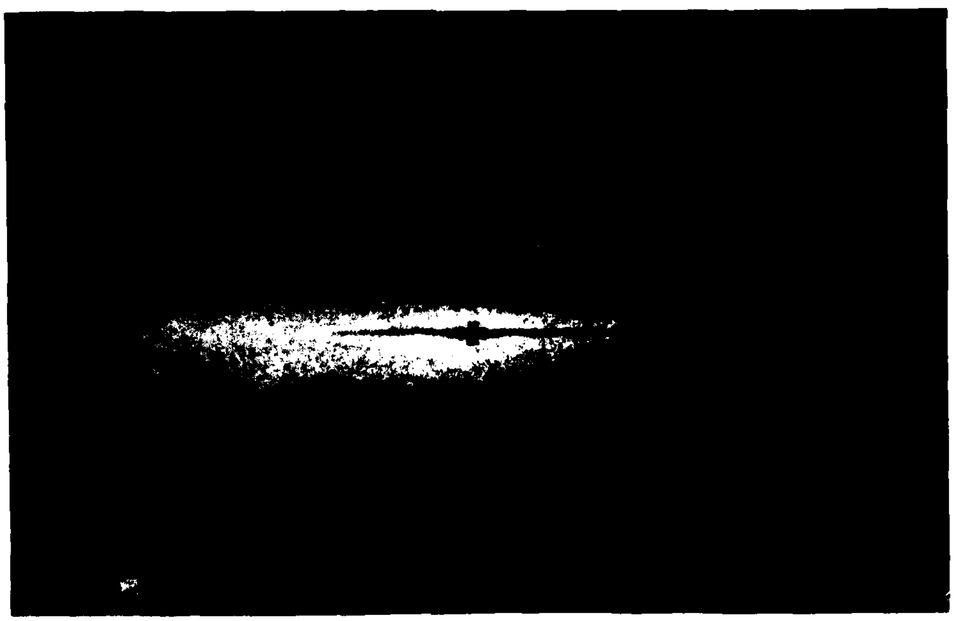


Figure 1-1. Milky Way galaxy (sideview).

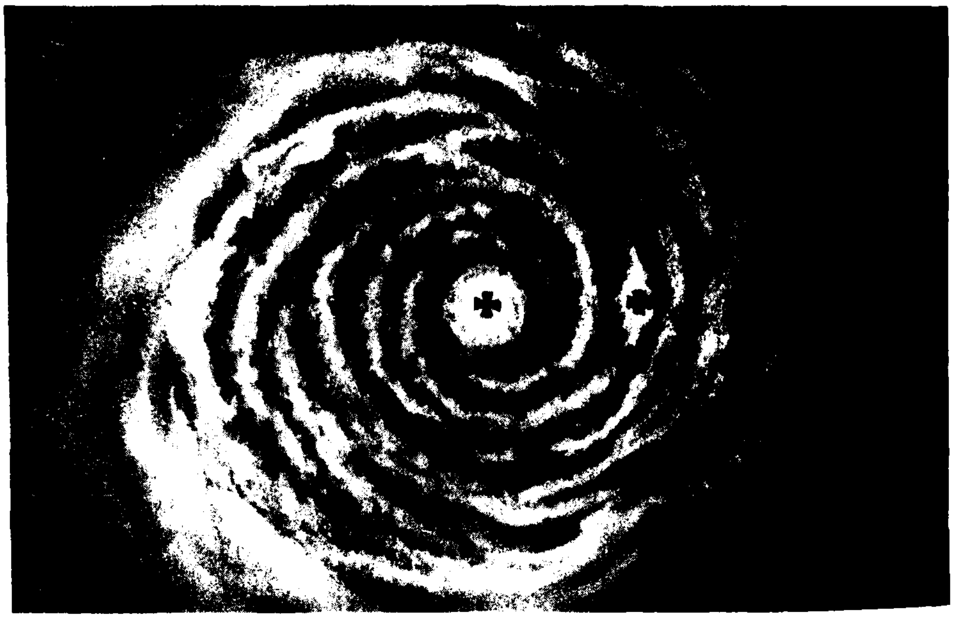


Figure 1-2. Milky Way galaxy (top view).

Sir James Jeans has provided an analogy which helps. He suggests that the number of stars in the universe is something like *all* the grains of sand on *all* the beaches of *all* the earth.

To complete the modern idea of this stupendous structure, scientists believe the whole thing is expanding. Every galaxy is rushing away from every other galaxy at tremendous speed! The most distant galaxy visible in a telescope is racing away at about 75,000 miles per second or 270 million miles per hour.

THE MILITARY SPACE ENVIRONMENT

Most of the space discussed above has little military importance today. However, the near-earth space is important for many military operations. Its characteristics set boundary conditions on both the operations conducted and the equipment required. We will devote the remainder of this chapter to the hazards and physical conditions of near-earth space. We will not define "near-earth" precisely, but we understand it to mean not more than 100 million miles from earth. The sun is approximately 93 million miles away. Thus the sun, the inner planets, and the space between them is the concern of this section. Emphasis will be upon this side of the moon.

The beginning of space is the first topic for consideration. It does not begin at the surface of the earth because that is where the atmosphere begins. At an altitude of 10,000 feet the oxygen pressure of the atmosphere is not great enough to keep people efficient over a long period of time. Many people become acclimatized to altitudes of 10,000 feet and higher. But for a person who lives near sea level, the oxygen pressure at levels above 10,000 feet is insufficient to sustain active and efficient performance. Thus, the Air Force requires the use of supplemental oxygen by crew members at altitudes above 10,000 feet.

Approximately one-half of the earth's atmosphere is below an altitude of three miles. At an altitude of approximately nine miles supplemental oxygen fails as a sufficient aid to sustain human life. Here the combined pressure of carbon dioxide and water vapor in the lungs equals the outside atmosphere pressure and breathing cannot take place without supplemental pressure. At this altitude pressure cabins or pressure suits become a necessity.

The vapor pressure of a person's body fluids is about 47 millimeters of mercury. As soon as the atmospheric pressure drops to this level, bubbles of water vapor and other gases appear in the body fluids. This means literally that the blood will boil. The gas bubbles first appear on the mucous membranes of the mouth and eyes and later in the veins and arteries. This would happen to an unprotected person at an altitude of 12 miles. However, supplemental pressure will suppress this evil.

At 15 miles compressing the outside air to pressurize a cabin is no longer effective. At that altitude the air density is about one twenty-seventh of the sea level value. Compressing this thin air is not an impossible task but it certainly is an uneconomical task. Further, the act of compressing air would involve undesirable heat transfer to the air. Finally, at this altitude the atmosphere contains a significant percentage of ozone. If the aircraft compresses the *ozone* it would poison the cabin atmosphere. Above this altitude the cabin or space suit must have a supply of both oxygen and pressure independent of the outside atmosphere. As far as people are concerned, space begins 15 miles above the earth's surface. Above this altitude they must take everything they need with them. The environment will supply them with neither food nor air. They need a sealed environment containing necessary supplies from earth.

Five miles further out, at the 20-mile level, is the operating limit for turbojet engines. At 28 miles ramjet engines do not have enough air to operate. Above this altitude the craft must supply the engines with both a fuel and an oxidizer. Thus, to a propulsion engineer, 28 miles above the earth is the beginning of space. Above this, the engineer must use rockets. In one sense space begins at 50 miles. Flight above this altitude earns a crew member the right to wear astronaut's wings.

In 1964, a New York law firm asked the Air Force Office of Aerospace Research to define the beginning of space. This scientific organization based the answer on aerodynamic forces. Usually, aeronautical engineers can neglect such forces acting on ballistic reentry vehicles, lifting reentry vehicles, and boost-glide orbital vehicles at altitudes above 100 kilometers or 62 miles. Thus for the aeronautical engineer concerned with lift and drag, space may begin at 62 miles.

At approximately 100 miles above the earth is a region of darkness and utter silence. This is the region of the black sky. The stars appear as brilliant points of light, and between them is absolute black because there is not enough air to scatter light. Neither is there enough air to carry sound or shock waves. There are no sonic booms.

Another reference frame for the question of the boundary of outer space is that of international law. Lt Col Robert H. Farris did a detailed study of this matter. He points out that the three treaties on space avoid an explicit altitude that defines the limit of space. However, in conventional and customary law, the major space powers accept "the lowest perigee attained by orbiting space vehicles as the present lower boundary of outer space". A potential challenge to this acceptance occurred in December 1976 when eight countries on the equator issued declarations of sovereignty over the synchronous orbit positions above their countries. Further, Colombia, Uganda, Brazil, Ecuador, Congo People's Republic, Indonesia, Kenya, and Zaire stated they would defend such areas. As of 1980 the major nations of the world consider these claims to be null and void under international law. These nations consider space to be international territory.

From the above discussion it is clear that there are many answers to the question, "Where does space begin?" The acceptable answer depends upon the reference frame in which the question is asked.

"Is space really nothing?" The answer is "no". Surprising amounts of matter fill it and energy floods it. Consider the density of matter by starting up from the earth's surface. At the surface the concentration of particles in air is about a million, million, million particles per cubic centimeter ($10^{18}/cm^3$). There is a decrease in particle density with altitudes and the *average* figures given are only approximate. An average figure for the zone between 7 miles and 50 miles is about $10^{14}/cm^3$. From 50 miles to 600 miles an average figure is about a million particles per cubic centimeter ($10^6/cm^3$). From 600 miles to 1200 miles there are still approximately 100 particles per cubic centimeter. We will find approximately one particle per cubic centimeter above 1,200 miles. This is certainly far from nothing. There are localized conditions that cause a much higher particle density. We will discuss some of these conditions later. Electromagnetic energy in many forms floods all of space. This energy comes from the sun, the stars in the Milky Way, and even other galaxies in the universe.

From another point of view, space is "not much." Consider air pressure. At the earth's surface the pressure varies around 760 millimeters of mercury. Above 1,200 miles the pressure is much less than one millimeter of mercury. It probably is around 10^{-12} to 10^{-16} Torr.* This pressure is so low that scientists often call it a "hard vacuum." It causes some unexpected phenomena. We will discuss a few of these later as illustrations.

In the atmosphere at least a single layer of absorbed gas covers any metal. In a hard vacuum this film of gas bleeds into space. Metals touching each other tend to weld together. In the atmosphere this doesn't happen because the thin film of air acts as a lubricant keeping the metals apart. To prevent this "cold welding" in space, scientists must take special measures.

Some metals are stronger in a hard vacuum. If a crack forms in a metallic surface when air surrounds the metal, molecules of air immediately enter the crack. Chemical reaction with the metal occurs. If the reaction product is more voluminous than the original crack, a wedging action occurs and enlarges the crack. In a hard vacuum a chemical reaction causing an enlargement of a crack does not occur. Thus some metals may be stronger in space than they are on earth.

To study space effects such as the above, it is useful to simulate a hard vacuum on earth. Late in 1965 the Air Force completed a large aerospace environmental simulation chamber at the Arnold Engineering Development Center in Tennessee.

Electromagnetic Radiation

Visible light, ultra-violet, X rays, infrared, radio, and other forms of energy can travel through the hard vacuum of space as electromagnetic radiation. This term refers to the fact that radiation consists of a varying electric and a varying magnetic field. Together these fields form a wave. It is possible to transmit such a wave through a vacuum. This transmission does not require the presence of a material medium. This form of energy floods all of space. Its intensity varies with proximity to the sun, or to a star.

Notice in figure 1-3 that visible light covers only a very narrow band in the spectrum of electromagnetic energy. The entire spectrum ranges from frequencies of about 10^4 Hertz up to about 10^{24} Hertz.

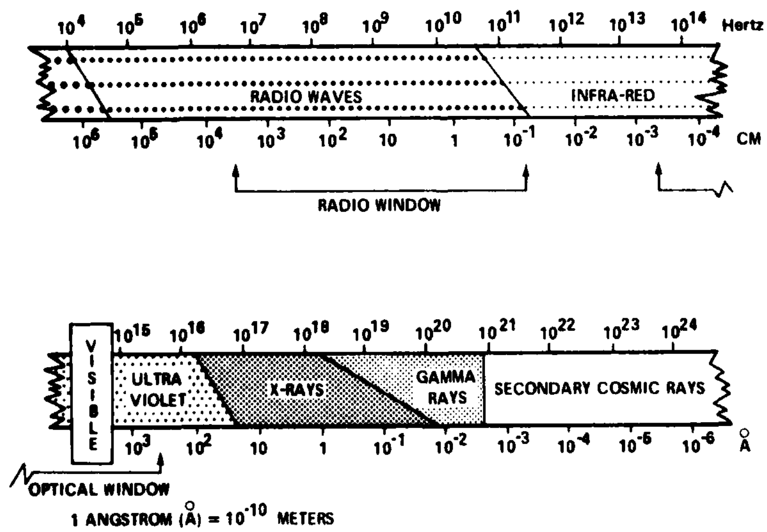


Figure 1-3. Spectrum of electromagnetic energy

*A Torr is the same as one millimeter of mercury. It is named in honor of Torricelli and is commonly used in low pressure measurements.

Both people and machines have limits of heat and cold they can withstand for any given period of time. In space, the temperature of an unprotected object will rise rapidly on the sunlit side to unacceptable levels, while simultaneously dropping on the shaded side. Various types of materials to reflect sunlight and insulate the vehicle or astronaut are necessary to maintain acceptable temperatures. Further, equipment generates heat that the spacecraft must reject or dump by some means. In a vacuum, radiating the excess heat into space is the only way to do this. The thermal design of spacecraft, tankage, space suits—in fact, the entire system—must take the heat balance into careful consideration.

Meteoroids and Micrometeoroids

The three terms "meteoroid," "meteor," and "meteorite" have similar meaning and often people use them interchangeably. Meteroid refers to a particle, large or small, moving in space. When a meteoroid enters the atmosphere and begins to glow, we call it a meteor. If that same particle survives the trip through the atmosphere and hits the earth, we call the remnant a meteorite. Some meteoroids must be very large because people have found some with masses of several tons. Most meteoroids are quite small. We call these extremely small meteoroids micrometeoroids.

Meteoroids and micrometeoroids move with speeds varying from about 30,000 miles per hour to 160,000 miles per hour. At these speeds, impact between a satellite and a large meteoroid would be catastrophic. Impacts between micrometeoroids and a satellite would not be catastrophic but could erode the satellite's surface. This erosion is a potentially serious hazard to optical surfaces, or to the material surfaces used for temperature control and heat exchange.

The meteorite material in orbit around the Sun in the vicinity of Earth's orbit consists of rocky or iron particles, with densities of about 3 and 8 gm/cm^3 respectivelv. Many of these particles combine into clusters or small sections. We call the remainder the meteoric background or sporadic meteorites. Measurements indicate that the density of dust particles near the earth is higher than the interplanetary background by a significant amount. This material could be dust particles separated from the moon, or interplanetary particles captured by earth's gravity.

Mysteriously, several satellites have ceased to function, and there is some conjecture that meteoroid impact damage may have been the cause. Scientists have studied extensively the probability of meteoroid and micrometeoroid impact. They have employed many methods. They have captured micrometeoroids so small that it would take about 125 of them to equal the thickness of a piece of paper and have studied them with an electron microscope. The Pegasus satellites, Explorer XXIII and Explorer XVI, are examples of satellites used to study the problem. The Pegasus satellites reported the number of penetrations of their panel materials. Aluminum was one type of panel material.

Explorer XVI, launched December 16, 1962, reported 62 meteoroid penetrations in 7½ months of space travel. Over 150 micrometeoroids hit Mariner IV on its trip to Mars. These and other data have led to the conclusion that the probability of a satellite being hit by very small micrometeoroids is high and of being hit by catastrophically large ones is small.

Figure 1-4 provides some idea of the hit probabilities involved in the Apollo project. The basis for the probabilities is an estimated vehicle cross section of 10 square meters and an exposure of 10 days. The horizontal axis shows mass of the meteoroids or micrometeoroids in grams. About five grams equals the weight of a nickel. The vertical axis is probability scaled until the axis reaches the figure one. This figure means that a meteoroid or micrometeoroid is certain to hit the vehicle. Above that, the scale means number of hits in a 10-day period. Notice that the probability of collision with a particle that weighs 10^{-5} grams is 1.0. This kind of particle would penetrate about 0.05 cm into an aluminum skin.

In general, the problem of meteoroid hazard is not as great as was thought, but it is not negligible. Space suits and capsules provide protection. The possibility of a catastrophic hit remains; therefore, scientists are continuing extensive studies.

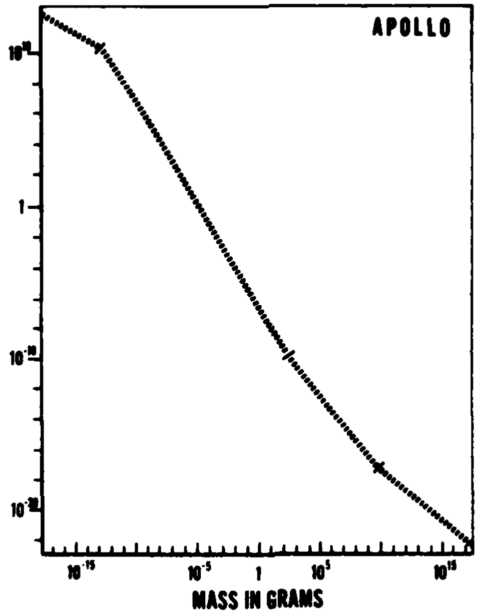


Figure 1-4. Meteorite probabilities in Apollo project.

Solar Activity

The sun, which radiates both electromagnetic (EM) and corpuscular streams, determines to a large extent the space environment. Many phenomena on earth are closely related to solar activity, although the relationship is not well understood.

The sun is radiating energy constantly even under quiet conditions. We call this the solar wind. Solar flares of varying intensity that appear suddenly over a small area punctuate this quiet at irregular intervals. These flares range from weak ones that last only 5 to 10 minutes, to strong ones that last several hours. Sporadic emission of EM energy (ultra-violet, X-ray, and radio frequencies) and the ejection of particles (mostly protons and electrons) up to high-energy solar cosmic rays accompany the flares. The total energy of solar cosmic rays ejected by a powerful flare may reach a billion-billion (10^{18}) kw-hrs. The Air Force Geophysics Laboratory (AFGL) developed an objective method of forecasting solar flare activity in FY 1981. The basis of this method is a quantitative approach to flare predictability. Scientists have proven this method is statistically accurate.

Cycles of various lengths characterize solar activity. The 11-year cycle measures the number of sunspots. There is a complex and poorly understood association between the sunspots and flares. A change in the sign of the overall solar magnetic field characterizes the 22-year magnetic cycle. Apparently this governs the index of the frequency of flares. The 80-year cycle is a quasi-periodic change in the maxima of the 11-year cycles. The most recent minimum occurred in 1980.

Electromagnetic radiation of the sun. The spectrum of solar EM radiation extends from the radio range to the X-ray frequencies, expanding to slightly higher frequencies during flares. We call the total energy incident on the earth's atmosphere the solar constant. Scientists know this amount very well, being about 1.36×10^6 erg/cm²-sec, or 1.95 cal/cm²-min. The atmosphere has two "windows" that are transparent to wavelengths from about 3,000Å to 40,000Å (the "optical window"), and from a few millimeters to about 15 meters (the "radio window") (see fig. 1-3).

Although the energy per photon in the region of shortwave radiation (10^{14} Hz and higher) is low

compared to the corpuscular ionizing radiation, the total energy transport is much higher due to the greater flux. This radiation can be very harmful to an unprotected person, and can change the surface properties of various materials.

The radiation received from the sun is constantly changing due to the sun's 27-day rotation period. Since the sun is not a solid, the period of the rotation is really an average. The sun rotates faster near its equator than at higher latitudes, and the rotation varies with time. We do not know the reasons for the changes in rotation rates.

The radio emission of the sun is quite complex. The sun emits three general classes of energy in this frequency range. First, there is a constant background over the whole radio spectrum from the "quiet" sun. Second, there is a slowly changing component related to sunspots. Third, there is sporadic emission related to centers of activity, such as flares.

Corpuscular radiation of the sun. We can divide the corpuscular radiation of the sun into constant, continuous emission of particles—the solar wind—and sporadic expulsion of intensive streams of plasma and charged particles—the corpuscular streams and solar cosmic rays. This division is arbitrary, reflecting the dependence on time of these types of radiation. It emphasizes the constant existence of the solar wind and variations in its velocity and density that never fall below the minimum values of 250km/sec and 0.5 particles/cm^3 at the level of the earth's orbit. Scientists generally consider separately, as solar corpuscular streams, the stronger streams of solar plasma observed sporadically, that is, the reinforced streams of the solar wind. They introduced this concept before the discovery of the solar wind to explain various geophysical phenomena that correlated with certain phenomena on the sun. The solar corpuscular streams reach velocities of about 1,600 km/sec at particle densities up to 100/cm^3. After their formation, these intense streams move through the quiet, slow portions of the solar wind, disrupt the stable structure of interplanetary space, and cause various disturbances. The solar wind and corpuscular streams are the most important components of solar corpuscular radiation, and determine conditions in interplanetary space.

Solar corpuscular radiation causes the sun to lose an average of a million tons of matter per second, which corresponds to 10^{-22} solar masses per second, assuming spherical symmetry of the solar wind. However, other published data indicate that the solar wind is not spherically symmetrical.

Another type of corpuscular radiation, solar cosmic rays, consists of high-energy charged particles (from 30 to 50 keV/nucleus to several GeV/nucleus). Recent studies indicate that every bright chromospheric flare on the sun probably generates solar cosmic rays.

Solar cosmic rays are apt tools for the study of interplanetary space. They illuminate the solar system and allow determination of its various characteristics. After large solar flares, great fluxes of solar cosmic rays sometimes represent a serious radiation danger for space flight.

Solar wind. Because of the high temperature of the sun's corona, protons and electrons beyond a certain distance from the sun acquire velocities in excess of the escape velocity from the sun. Thus there is a continuous outward flow of charged particles in all directions from the sun. We call this the solar wind. It is a plasma wind, rather than a gas wind. Its velocity and density vary with sunspot activity. During the time of sunspot minimum at the Earth's distance from the Sun, the density is about 100 particles per cubic centimeter. The speed is about 300 miles per second. At sunspot maximum the corresponding density is approximately 10,000 particles per cubic centimeter, and the speed is about 900 miles per second. When the solar wind encounters the earth's magnetosphere, it flows around the magnetosphere, which becomes flattened in the process. On the dark side of the earth, the magnetosphere becomes elongated.

The energy of the particles in the solar wind is not high. We do not expect any hazard to the people on earth from this wind. However, it is either the cause of or a contributor to, the aurora that illuminates the polar night sky. Within the past few years, scientists have combined ground-based observations with information that rockets and earth satellites acquired to provide an explanation of the previously baffling beauty. The magnetosphere of the earth acts like a giant, cathode ray tube, directing charged particles from the solar wind into beams and focusing them on the earth's polar

regions. The aurora is a fluorescent luminosity produced by the electrons and protons of the solar wind which strike atoms and molecules of oxygen and nitrogen high in the atmosphere of the polar region.

Solar flares. The high-speed solar protons emitted by a solar flare are probably the most potent of the radiation hazards to space flight. Flares themselves are the most spectacular disturbances seen on the sun. We can observe them optically as a sudden, large increase in light from a portion of the sun's atmosphere. A flare may spread in area during its lifetime, which may be from several minutes to a few hours. We classify flares according to a range of importance from zero to four. There is a relationship between the number of sunspots and the frequency of flare formation, but the most important flares do not necessarily occur at sunspot maximum.

There are many events that occur on earth following a solar flare although not many flares produce *all* of the possible events. In addition to the increase in visible light, minutes after the start of a flare there is a sudden ionospheric disturbance (SID) in the earth's ionosphere. This SID causes shortwave fadeout, resulting in the loss of long-range communications for 15 minutes to 1 hour. X-rays emitted by the flare probably cause the SIDs. During the first few minutes of a flare there may be a radio noise storm, consisting of bursts of noise over a wide range of frequencies. In addition, there may be disturbances in the earth's magnetic field, changes in the auroras, and *decreases* in galactic cosmic ray intensity. However, from the point of view of a space traveller, by far the most important effect is the marked increase in solar protons. The energy of these protons ranges from approximately 10 million electron volts to approximately 500 million electron volts. The flux may be quite high. Consequently, the dose of radiation accumulated during exposure to the solar protons may vary from negligible to well above a lethal dose.

The Van Allen Radiation Belt

Another problem that people must overcome in venturing into space is trapped radiation. As a result of experiments conducted in 1958 with the US Explorer I satellite and subsequent experiments, Dr James A. Van Allen and his associates discovered the existence of geomagnetically-trapped particles encircling the earth. When electrons and protons, and perhaps some other charged particles, encounter the earth's magnetic field, the field traps many of them. They oscillate back and forth along the lines of force, and since the magnetic field completely encircles the earth, the trapped particles completely encircle the earth.

The belt has an inner and outer portion. Recent data show that both protons and electrons permeate the toroidal-shaped volume occupied by the Van Allen radiation. The protons are most intense at approximately 2,200 miles. The electron flux peaks at approximately 9,900 miles. Often we call the low-particle density separating the two belts, the slot. In this particular volume of space some phenomena, as yet not fully understood, reduces the lifetime of the charged particles.

Table 1-1.
Solar Flare Classification

Importance	*Area* (Millionths of Solar Hemisphere)*	*Average Duration*
0	Less than 100	17 min
1	100 to 249	32 min
2	250 to 599	69 min
3	600 to 1200	145 min
4	Greater than 1200	145 min
BRIGHTNESS CATEGORIES	FAINT (F) NORMAL (N) BRILLIANT (B)	

* The area of the earth's disk is approximately equivalent to the area of an Importance one flare.

The inner Van Allen belt starts at an altitude of approximately 250 miles to 750 miles, depending upon the latitude. It extends within the vicinity of 6,200 miles where it begins to overlap the outer belt and where the slot begins. The inner belt extends from 45 degrees north latitude to about 45 degrees south latitude.

The outer Van Allen belt begins around 6,200 miles and extends to an altitude that varies from 37,000 to 52,000 miles. The upper boundary is dependent upon the activity of the sun.

A July 1962 high-altitude nuclear device test affected both the inner and outer belts. Radiation in both belts increased after detonation. The test eliminated for a long time the low radiation slot separating the two belts.

Experience has shown that space vehicles in low circular orbit (125-350 miles) receive an insignificant amount of radiation from the Van Allen zones. However, a vehicle in a highly eccentric orbit or one in a high altitude circular orbit can receive an important dose. For example, a satellite in a synchronous orbit over the equator will be close enough to the center of the outer zone to accumulate a hazardous dose. But, as demonstrated by the Apollo lunar missions, people can transit safely these zones in a spacecraft with minimal shielding by judicious selection of the flight trajectory.

The Van Allen belt varies daily with changes in the magnetosphere. On the sun side of earth it is flattened. On the night side of earth it is elongated.

Cosmic Rays

Cosmic rays are high-energy corpuscular radiation. They originate from two sources, either the sun or outside the solar system. The composition of this radiation consists of protons and electrons (about 95 percent), some helium nuclei (about 5 percent), and other nuclei ranging up to very heavy particles (less than 1 percent). The galactic cosmic rays can be extremely energetic, but do not pose a serious threat due to the low flux. The solar cosmic rays (the high-energy portion of the sun's corpuscular radiation) are not a serious threat to people except during periods of flare activity. Then the radiation can increase a thousandfold over short periods of time. Manned operations during such conditions would require heavy on-board shielding, which is generally impractical today. The only alternative would be to curtail the mission.

Radiation Hazard Summary

We can summarize the radiation hazard of space as follows:

1. Shielding of electromagnetic energy in space is possible. Thus, electromagnetic energy is not a serious threat to life.
2. Electromagnetic energy may upset radio communication and guidance equipment.
3. Galactic cosmic rays and solar wind do not present a serious threat to space travel.
4. Van Allen radiation does present a serious threat but we know the location of the belts well enough that we can plan flight trajectories to limit time spent in the hazardous regions.
5. Protons emitted at the time of a solar flare present the greatest uncertainty and the greatest threat to manned flight in regions beyond the protection of the earth's atmosphere and magnetosphere.

MILITARY SPACE RESEARCH

The space environment affects Air Force systems that operate in the near-earth space environment. Near-earth space is a varying environment, and the field and particles of its composition vary diurnally, seasonally, and, in particular, in response to activity on the sun. Such variations or disturbances disrupt and degrade communication systems, surveillance systems, detection and tracking systems, and interfere with the operation of electronic devices and detectors on satellites. A number of laboratories within the Air Force Geophysics Laboratory (AFGL) (formerly Air

Force Cambridge Research Laboratories) are engaged deeply in studying the space environment and understanding the underlying phenomena.

The Space Physics Laboratory conducts research on the near-earth space environment to understand the behavior of that environment and to define its parameters for the purpose of providing satellite design criteria, and of developing a capability to predict disruptive disturbances. Thus, this laboratory's research program studies the varying magnetic and electric fields in the magnetosphere, and the particle fluxes and distributions occurring in that region of space in which Air Force satellites operate. The research involves experimental and theoretical efforts. Ground-based observations and instrumentation on rockets and satellites provide the data. Scientists applied considerable theoretical effort to the development of a model of the magnetosphere that the Air Force systems can use to cope with the real problems they experience operating in or using the near-earth space environment.

Scientists detected the geophysical effects of solar activity about a century ago when they observed variations in geomagnetic storms paralleling variations in solar activity. The telescopes and magnetometers required for this discovery were relatively simple tools by modern standards. As the technology of observation advanced, more and more features of the solar-terrestrial relationship became apparent. The invention of the optical spectroheliograph and the development, in the 1920s, of radio communications revealed the tendency of large flares, "chromospheric eruptions," to cause radio blackouts. More recently, space science confirmed the existence of the solar wind, the magnetosphere, and the geomagnetic effects of their interaction. A particularly important discovery was that fast streams of solar wind particles originated from vast voids in the corona that scientists promptly named "coronal holes." Technological progress has advanced our knowledge, but has exposed us to more and more practical problems due to the geophysical responses to solar influences. Radio blackouts, and more recently, variations in the density of the upper atmosphere have become important, and, as the technology continues to advance, other Air Force activities will encounter other environmental factors related to solar activity.

The link between solar disturbances and the earth are the X rays and energetic particles emitted by solar flares and flare-related phenomena, and the relatively low-energy particles of the solar wind and the embedded interplanetary magnetic field. Sporadic variations in these radiations produce changes in the ionosphere and magnetosphere. The operational problems come from these.

In extreme instances surges on hard-wire communication that completely garble messages accompany degradation or complete blackout of long distance radio propagation. Severe clutter on over-the-horizon (OTH) radar systems either drowns the signals or produces signatures difficult to distinguish from those of targeting and surveillance position errors, and grossly affects predicted and planned reentry times and coordinates. Temporary blackouts of satellite surveillance equipment due to showers of energetic particles last for several hours in the most extreme cases. Bursts of X rays and particles can confuse space monitors of surreptitious nuclear activity. Finally, the evidence for the influence of solar variations on global weather patterns is now fairly definite, although the nature of the interaction is still a complete mystery.

Energetic Particle Research

Energetic particle fluxes limit space operations. By depositing energy in the earth's atmosphere, ionizing and heating the ambient gases, particularly at high latitudes, they degrade, and sometimes inhibit, the operation of systems that rely on long-distance propagation of electromagnetic radiation. The increased Air Force utilization of spaceborne sensor and communication systems has demonstrated requirements for better knowledge of particle data at all altitudes. Low-orbit systems are affected by high atmosphere density changes. These changes partially result from heat transfer to the atmosphere as energetic particles are stopped. At synchronous altitude, the mean free path of particles is so long that electric charging due to particle impact permits sizeable charge buildup on satellite surfaces, producing electrical interference.

Both electric and magnetic fields affect the motion of charged particles. Magnetic fields can affect only the direction of motion and thus deflect particles from some volumes and concentrate them in others, while electric fields can modify their energy as well. The relatively stable terrestrial magnetic field consequently tends to exclude low energy inflowing particles from equatorial regions and to contain those particles already in orbit near the earth in a restricted region, that is, the so-called Van Allen belts.

High solar activity not only brings solar produced particles to the earth but also changes the electromagnetic field configuration, which modifies the resident particle populations. Systems designed to operate properly in the average particle environment experience malfunctions or anomalous behavior during and after extreme solar activity. The malfunction of Defense Satellite Communications System (DSCS) satellites occurred predominately during magnetic substorms. The influx of plasma from the earth's magnetotail to synchronous altitude apparently caused the malfunctions.

AFGL's program of solar energetic particle study includes both direct acquisition of solar particle data on USAF satellite-borne sensors and its subsequent analysis and theoretical model studies of solar particle propagation in the interplanetary space and in the magnetosphere. This program develops models of solar proton events, particularly those with earth-expected intensity. The Global Weather Central-Air Weather Service (GWC-AWS) uses these models to predict the intensity of particles as a function of time from observables available to GWC either from satellite-borne or ground-based sensors.

An example of such an event occurred on 13 14 January 1967, when a large flare erupted on the sun. The passage of the storm front caused severe distortion of the earth's magnetic field. Five satellites VELA 5 and 6, OGO 3, ATS 1, and Explorer 33 measured the event rather completely. Analysis revealed that the magnetosphere (and outer Van Allen belt) were depressed below the geosynchronous altitude of 22,225 statute miles.

Another solar event occurred October 30, 1972. Fortunately, during this period appropriate instruments were operating in two military satellites. One of these, OV5-6, was observing outside the magnetosphere, while the other, S72-1, was taking data at lower altitude above the polar caps and within the magnetosphere. These data permitted study of the transmission characteristics of particles from outside to inside of the magnetosphere.

Scientists used the S72-1 satellite to continue studies of high-energy trapped proton fluxes in the South Atlantic anomaly while the spacecraft traversed the lower latitudes. The omnidirectional fluxes averaged over the period October 1972 to February 1973 were in agreement with nuclear emulsion data taken in 1961 62 before the "Starfish" high altitude test. Since this time interval of 11 years is coincident with a solar cycle, this further supports earlier conclusions that the atmosphere controls particle lifetimes.

The Air Force is phasing out the VELA satellite system. To fill the gap in the data base required by the military, AFGL incorporated the appropriate particle instruments into the SOLRAD-HI satellite system, which became operational in 1976. It consists of two satellites located outside the magnetosphere at 19 earth radii, together with a dedicated readout station operated by the US Navy. The Navy sends these data to Air Weather Service Global Weather Central (AWS/GWC) in real time for operational use.

Research on prediction of solar particle events continues with the objective of improving the accuracy of the proton prediction system originally developed by the Space Physics Laboratory and currently used by the Air Weather Service. One of the problems addressed was the updating of initial prediction parameters from real-time satellite data to get a better answer to the question: "Once we identify a solar event and it has realized the maximum flux, when will the event end?" (That is, when will the enhanced intensity recede to a background level?). This study showed that the maximum flux, which occurs early in a solar particle event, determines the equilibrium condition from which scientists can predict the future decay rate and the end of the event. In practice, scientists are updating the predictions continually as additional time data become available.

Spacecraft Charging

AFGL participated in a study to establish the cause of anomalous behavior of the DSCS at times of geophysical activity. As a result of this study, scientists have linked spacecraft charging at synchronous orbit to the transport of high-energy plasma particles from the magnetospheric tail to the synchronous regime. The charging process is a result of the complex interaction between a satellite and its environment. During a magnetospheric substorm, a satellite, wrapped in its thermal blanket, acts much like a capacitor immersed in a high temperature plasma.

Because of the anomalous behaviour of military spacecraft at synchronous orbit, the Air Force initiated a research satellite to investigate the phenomenon of spacecraft charging at high altitude (SCATHA). Environmental monitors aboard the SCATHA satellite provided data during the magnetospheric quiet and storm times. Engineering experiments detected and analyzed in detail the electrical charging phenomena. The Air Force correlated scientific data and engineering information to relate cause and effect in satellite charging. The SCATHA program concluded in FY 1981. Its final contributions were to demonstrate the concept of active discharging as exemplified by a plasma gun used for discharging the spacecraft and a complete code for calculating satellite surface potentials. SCATHA developed a method for modeling active charging/discharging and created an atlas specifying the geosynchronous environment.

Geomagnetism

Magnetic activity levels control the maximum and minimum usable frequencies of a high frequency communications network, and advance assignment of a given channel requires a knowledge of the expected activity levels. The communications personnel must operate the aeromagnetic detection, perimeter safeguards, and other magnetic devices at reduced sensitivity when their detection bandwidth corresponds to the natural frequency components of magnetic disturbances. Another example of the effects of geomagnetic activity is the variation in atmospheric drag experienced by artificial satellites that relates to enhanced levels of activity. As a consequence, this atmospheric drag can change the orbital parameters of the satellites significantly, thus changing the predicted position of the satellite as a function of time.

We can divide geomagnetism research into two parts: measurements, both ground-based and space; and theory and analysis. Scientists make particles and field measurements using suitably instrumented rockets launched from Fort Churchill, Canada, during magnetic storms. A magnetometer network within the United States will provide the ground measurements. Scientists construct theoretical models of geomagnetic phenomena to aid in the interpretation of the data. They develop advanced methods of data assimilation, especially high resolution spectral techniques, and apply them to the analysis of the observations.

Space Forecasting

AFGL initiated the Space Forecasting program in the mid-1960s to develop techniques to observe, specify, and predict solar related changes in the earth's upper atmosphere and near-space environment that impact on Air Force operations. This interdisciplinary program draws on most of [illegible] environmental sciences: radio and optical solar astronomy, astrophysics, radio propagation [illegible] and the physics and chemistry of the upper atmosphere. Its very nature requires participation [illegible] in a variety of technical fields, such as energetic particles, ionospheric physics, solar [illegible] geomagnetism, and atmospheric density.

[illegible] inception, the Space Forecasting program has focused strongly on the operational [illegible] changing space environmental conditions. The Air Force established the Space [illegible] Group in 1966 to provide a convenient mechanism for coordinating the [illegible] scientists. This group, which meets three to four times a year, provides an ideal [illegible] technical information, for coupling research products with operational [illegible] directing and coordinating the efforts of in-house scientists and contractors

working in a broad spectrum of related scientific disciplines. In addition to the AFGL task scientists, representatives from Air Weather Service, National Oceanic and Atmospheric Administration (NOAA), Naval Research Laboratory (NRL), Electronic Systems Division (ESD), and Space Division of the Air Force Systems Command are periodically invited to participate at these meetings to coordinate related activities within their organizations.

Space forecasters face the task of predicting the occurrence of bursts of electromagnetic radiation and energetic particles from the sun, and specifying the effects of these bursts on the earth's atmosphere. These effects include absorption of radio waves in the ionosphere, geomagnetic disturbances, auroral activity, enhanced airglow levels, and atmospheric density variations.

Through the efforts of the Space Forecasting Coordination Group, the Air Force has identified specific problem areas related to existing and planned operational space environmental support requirements and initiated programs to solve them.

Solar Radio Astronomy Research

Research has shown that one of the best tools for studying solar activity is the radio emission. Apart from the spectrally narrow optical region, radio offers the only other spectral window for solar observations from the ground. Its quantitative data provide reliable predictors for geophysical phenomena. Continuing research provides an increasing number of associations, which help in understanding the mechanisms of the sun.

A balanced program of research satisfies operational needs by providing significant observational data and by investigating how the data are correlated with ionospheric and geophysical parameters. The solar radio astronomy research program at AFGL makes patrol-type (low angular resolution) measurements of the quiet sun and bursts, and high resolution measurements of active regions.

Absolute accuracy in the measurement of quiet-sun radio emission has become an operational necessity. The slowly varying component of solar radio emission relates well to atmospheric neutral density, which causes drag on space vehicles. Therefore, radio data provide an input to satellite position calculations. Recent results indicate that longer or shorter wavelength data may, at certain periods of the solar cycle, lead to better neutral-density predictions than the use of the traditional 10-cm wavelength data. The DOD widely uses the quiet-sun data for calibration of DOD telemetry-range antenna systems on a routine basis. For this application, the calibrators require day-to-day consistency and maintenance of this patrol data.

The Air Weather Service, through its Space Environmental Support System (SESS), disseminates the data. AFGL works closely with AWS in both operational and research applications. Sagamore Hill is the principal station of the Air Force solar radio network, continuously developing techniques and providing calibration for the other stations. A second generation solar radio network known as the Radio Solar Telescope Network (RSTN) is in the procurement stage, with AFGL responsible for engineering direction. Partly completed stations of the network at Manila, Athens, and Hawaii (all instrumented by AFGL) now complement the Sagamore Hill station.

BIBLIOGRAPHY

Air Force Cambridge Research Laboratories. *AFCRL in Space.* Bedford, Mass.: Office of Aerospace Research, 1967.

__________. *Report on Research at AFCRL, July 1972-1974.* AFCRL-TR-75-0288, May 1975.

Calvin, Melvin and Oleg G. Gazenko, ed. *Foundations of Space Biology and Medicine.* NASA Scientific and Technical Information Office, Washington, D.C.: US Government Printing Office, 1975.

Cosby, William A. and Robert G. Lyle. *The Meteoroid Environment and its Effects on Materials and Equipment.* Washington, D.C.: US Government Printing Office, 1965.

Farris, Robert H. *The Problem of Delimitation in Space Law.* PhD Diss., Notre Dame University, 1974.

Flinders, Dale J. "The Space Forecasting System Confluence of Military and Scientific Interests." *Air University Review* (November-December 1969): 37-50.

Fourth Weather Wing, 4WWM 105-1. *Forecasting Solar Activities and Geophysical Responses.* Solar Forecasting Facility Headquarters, Ent AFB, Colo., 28 June 1968.

Fourth Weather Wing, 4WWP 105-1, *Guide to Solar Geophysical Activity*. 4WW Headquarters, Ent AFB, Col., 1 February 1969.

Glasstone, Samuel. *Sourcebook on the Space Sciences*. New York: D. Van Nostrand Co., 1965.

Howard, Robert. "The Rotation of the Sun." *Scientific American*. (April 1975):106 114.

Office of Aerospace Research. "The Videometer, a New Instrument for Solar Flare Analysis." *OAR Research Review*. vol. 9, no. 1 (1970):7.

Parker, E.N. "The Sun." *Scientific American*. (September 1975):42 50.

Seaborn, H.T., ed. *TRW Space Log*. vol 8, no. 3 (Fall 1968).

"State Department Monitors Sovereignty Claim." *Aviation Week and Space Technology*. (20 December 1976):29.

Van Allen, James A. "Interplanetary Particles and Fields." *Scientific American*. (September 1975):160 173.

Zimmerman, Frank J. *Meteoroid Threat to Extravehicular Spacesuite Assemblies*. Aerospace Medical Research Laboratory, AFSC, Wright-Patterson AFB, Ohio, AMRL-TR-68-86, June 1969.

Chapter 2

ORBITAL MECHANICS

The study of trajectories and orbits of vehicles in space is not a new science but is the application of the concepts of celestial mechanics to space vehicles. Celestial mechanics, which is mainly concerned with the determination of trajectories and orbits in space, has been of interest to people for a long time. When the orbiting bodies are man-made (rather than celestial), the topic generally is known as orbital mechanics.

The early Greeks postulated a fixed earth with the planets and other celestial bodies moving around the earth, a geocentric universe. About 300 B.C., Aristarchus of Samos suggested that the sun was fixed and that the planets, including the earth, were in circular orbits around the sun. Because Aristarchus' ideas were too revolutionary for his day and age, the people rejected them, and the geocentric theory continued to be the accepted theory. In the second century A.D., Ptolemy amplified the geocentric theory by explaining the apparent motion of the planets by a "wheel inside a wheel" arrangement. According to this theory, the planets revolve about imaginary planets, which in turn revolve around the earth. It is surprising to note that, even though Ptolemy considered the system as geocentric, his calculations of the distance to the moon were in error by only 2 percent. Finally, in the year 1543, some 1800 years after Aristarchus had proposed a heliocentric (sun-centered) system, a Polish monk named Copernicus published his *De Revolutionibus Orbium Coelestium,* which again proposed the heliocentric theory. This work represented an advance, but there were still some inaccuracies in the theory. For example, Copernicus thought that the orbital paths of all planets were circles and that the centers of the circles were displaced from the center of the sun.

The next step in the field of celestial mechanics was a giant one made by a German astronomer, Johannes Kepler (1571-1630). After analyzing the data from his own observations and those of the Danish astronomer Tycho Brahe, Kepler stated his three laws of planetary motion.

A contemporary of Kepler's, named Galileo, proposed some new ideas and conducted experiments, the results of which finally caused acceptance of the heliocentric theory. Newton expanded and improved some of Galileo's ideas. These became the foundation for Newton's three laws of motion. Newton's laws of motion, with his law of universal gravitation, made it possible to prove mathematically that Kepler's laws of planetary motion are valid.

Kepler's and Newton's work brought celestial mechanics to its modern state of development, and the major improvements since the days of Newton have been mainly in mathematical techniques, which make orbital calculations easier.

Because the computation of orbits and trajectories is the basis for predicting and controlling the motion of all bodies in space, this chapter describes the fundamental principles of orbital mechanics, which is the basis for these computations. It shows how these principles apply to the orbits and trajectories used in space operations.

MOTION OF BODIES IN ORBIT

Bodies in space move in accordance with defined physical laws. We can analyze orbital paths by applying these laws to specific cases. Orbital motion is different from motion on the surface of the

earth. However, we can transfer many concepts and terms, and we can apply similar logic in both cases. An understanding of simplified linear and angular motion will permit a more thorough appreciation of a satellite's path in space.

Linear Motion

We observe bodies in space to be continuously in motion because they are in different positions at different times. In describing motion, it is important to use a reference system. Otherwise, misunderstanding and inaccuracies are likely to result. For example, a passenger on an airliner may say that the stewardess moves up the aisle at a rate of about 5 ft per sec, but, to the person on the ground, the stewardess moves at a rate of the aircraft's velocity plus 5 ft per sec. The person in the air and the person on the ground are not using the same reference system. Describing movement along a straight line, called rectilinear motion, will simplify the matter of a reference system for the present.

We can describe rectilinear motion in terms of speed, time, and distance. Speed is the distance traveled in a unit of time, or the time rate of change of distance. An object has uniform speed when it moves over equal distances in equal periods of time. Speed does not describe motion completely. We can describe motion more accurately if we give a direction as well as a speed. A speedometer tells how fast an automobile is going. If we associate a direction with speed, we can describe the motion as a velocity. A velocity has both a magnitude (speed) and a direction, and it is a vector quantity.

Uniform speed in a straight line is not the same as uniform speed along a curve. If a body has uniform motion along a straight line for a given time, the equation $v = \frac{s_f - s_o}{t_f - t_o}$ represents velocity. In the equation, s_o is the initial position, s_f is the final position, t_o is the initial time, and t_f is the final time. More simply, the velocity is the change in position divided by the change in time. The units of velocity are distance divided by time, such as feet per second or knots (nautical miles per hour).* Since velocity is a vector quantity, we may treat it mathematically or graphically as a vector.

If velocity is not constant from point to point (that is, if either direction or speed is changed), there is acceleration. Acceleration, which is also a vector quantity, is the time rate of change of velocity. The simplest type of acceleration is one in which the motion is always in the same direction and the velocity changes equal amounts in equal lengths of time. If this occurs, the acceleration is constant, and we can describe the motion as being uniformly accelerated.

The equation $a_{av} = \frac{v_f - v_o}{t_f - t_o}$ defines the average acceleration, over the specified time interval. A good example of a constant acceleration is that of a free-falling body in a vacuum near the surface of the earth. Scientists have measured this acceleration as approximately 32.2 ft per sec per sec, or 32.2 ft per sec^2. We usually give it the symbol g. Since an acceleration is a change in velocity over a period of time, its units are ft/sec^2, or more generally, a length over a time squared. Actually, constant acceleration rarely exists, but we can adapt the concepts of constant acceleration to situations where the acceleration is not constant.

The following three equations are useful in the solutions of problems involving linear motion:

$$1.\ s = v_o t + \frac{at^2}{2}$$

$$2.\ v_f = v_o + at$$

$$3.\ 2as = v_f^2 - v_o^2$$

where s is linear displacement, v_o is initial linear velocity, v_f is final linear velocity, a is constant linear acceleration, and t is the time interval.

*The nautical mile (NM) is one minute of a great circle and is approximately 6,080 feet or 1.15 statute miles.

Angular Motion

If a particle moves along the circumference of a circle with a constant tangential speed, the particle is in uniform circular motion. Since velocity signifies both speed and direction, however, the velocity is changing constantly because the direction of motion is changing constantly. We can define acceleration as the time rate of change of velocity. Since the velocity in uniform circular motion is changing, there must be an acceleration. If this acceleration acts in the direction of motion called the tangential direction, the magnitude of the velocity (the speed) would change. But, since the original statement assumed that the speed was constant, the acceleration in the tangential direction must be equal to zero. Therefore, any acceleration that exists must be perpendicular to the tangential direction, or any acceleration must be in the radial direction (along the radius).

Average speed is equal to the distance traveled divided by the elapsed time. For uniform circular motion, the distance in one lap around the circle is $2\pi r$ and is covered in one period (P). Period is the time required to make one trip around the circumference of the circle. Therefore, the tangential speed $v_t = \frac{2\pi r}{P}$. In uniform circular motion, the particle stays the same distance from the center, therefore, radial speed, $v_r = 0$. We have shown that $a_t = 0$; and we will show in the next section that $a_r = \frac{4\pi^2 r}{P^2} = \frac{v_t^2}{r}$.

Translatory motion is concerned with linear displacement, s; velocity, v; and acceleration, a. Angular motion uses an analogous set of quantities called angular displacement, θ; angular velocity, ω; and angular acceleration, α.

In describing angular motion, it is convenient to think of it in terms of the rotation of a radius arm (r), as figure 2-1 shows. The radius arm initially coincided with the polar axis, but later (t seconds), it was positioned as shown.

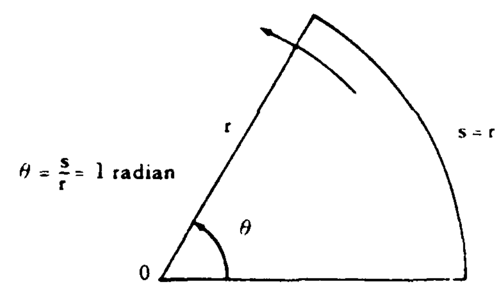


Figure 2-1. Position of radius arm as rotated one radian (57.3 degrees) from the starting point

We measure angular displacement (θ) in degrees or radians. A radian is the angle at the center of a circular arc which subtends an arc length equal to the radius length. If the length of s equaled the length of r, θ would be equal to one radian, or 57.3°. The central angle of a complete circle is 360° or 2π radians ($2\pi = 6.28$).

The following equations for angular motion are analogous to those studied earlier for rectilinear motion:

$$\theta = \frac{s}{r} \text{ radians}$$

$$\omega_{av} = \frac{\theta_1 - \theta_0}{t_1 - t_0} \text{ rad/sec}$$

$$\alpha_{av} = \frac{\omega_f - \omega_o}{t_f - t_o} \text{ rad/sec}^2$$

$$\omega_f = \omega_o + \alpha t$$

$$\theta = \omega_o t + \frac{\alpha t^2}{2}$$

$$2\alpha\theta = \omega_f^2 - \omega_o^2$$

In the equations, θ_f is final angular position; θ_o is initial angular position; s is linear displacement (arc length); r is the radius; ω is average angular speed; ω_f is final angular speed; ω_o is initial angular speed; t_f is final time; t_o is initial time; and α is constant angular acceleration.

If a body rotates about a center on a radius r, the tangential linear quantities relate to the angular quantities by the following formulas [where θ, ω, and α are in radians]:

$$s = r\theta$$

$$v_t = r\omega$$

$$a_t = r\alpha$$

LAWS OF MOTION

Natural bodies in space follow the basic laws of dynamics, as described by Newton's universal law of gravitation and his three laws of motion. By applying the basic laws and making use of calculus (also developed by Newton), we can explain and prove Kepler's three laws of planetary motion. It would be well to review Kepler's laws before stating Newton's law of universal gravitation, which is one of the laws upon which we base the computation of trajectories and orbits*, and Newton's three laws of motion, which describe terrestrial motion as well as celestial mechanics.

Kepler's Laws

From his observations and study, Kepler concluded that the planets travel around the sun in an orbit that is not quite circular. He stated his first law thus:

The orbit of each planet is an ellipse with the sun at one focus.

Later Newton found that he had to make certain refinements to Kepler's first law to take into account perturbing influences. As we apply the law to earthmade satellites, we must assume that perturbing influences like air resistance, the nonspherical (pear shape) shape of the earth, and the influence of other heavenly bodies are negligible. The law as applied to satellites is as follows: The orbit of a satellite is an ellipse with the center of the earth at one focus. The path of a ballistic missile, not including the powered and reentry portions, is also an ellipse, but one that happens to intersect the surface of the earth.

Kepler's second law, or law of areas, states:

Every planet revolves so that the line joining it to the center of the sun sweeps over equal areas in equal times.

To fit earth orbital systems, we should restate the law thus: Every satellite orbits so that the line joining it with the center of the earth sweeps over equal areas in equal time intervals.

*The terms "trajectory" and "orbit" are sometimes used interchangeably. Use of the term "trajectory" came to astronautics from ballistics, the science of the motion of projectiles shot from artillery or firearms, or of bombs dropped from aircraft. The term "orbit" is used in referring to natural bodies, spacecraft, and manmade satellites. It is the path made by a body in its revolution about another body, as by a planet about the sun or by an artificial satellite about the earth.

When the orbit is circular, the application of Kepler's second law is clear, as shown in figure 2-2. In making one complete revolution in a circular orbit, a satellite at a constant distance from the center of the earth (radius r) would sweep out eight equal areas in the total time period ($P = 1$). Each of these eight areas is equal and symmetrical. According to Kepler's second law, the time required to sweep out each of the eight areas is the same. When a satellite is traversing a circular orbit, therefore, its speed is constant.

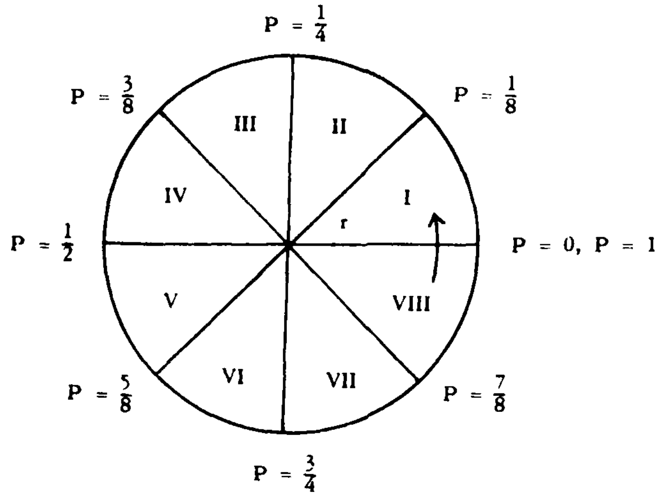


Figure 2-2. Law of Areas as applied to a circular orbit.

When the orbit is elliptical rather than circular, the application of Kepler's second law is not quite so easy to see; although the areas are equal, they are not symmetrical (fig. 2-3). Note that the arc of sector I is much longer than the arc of sector V. Therefore, since the radius vector sweeps equal areas in equal fractions of the total time period, the satellite must travel much faster around sector I (near perigee) than around sector V (near apogee). The perigee (a word derived from the Greek prefix *peri-*, meaning "near," and the Greek root *ge*, meaning "pertaining to the earth") is the point of the orbit nearest the earth. The apogee is that point in the orbit at the greatest distance from the earth (the Greek prefix *apo-* means "from" or "away from").

Kepler's third law, also known as the harmonic law, states:

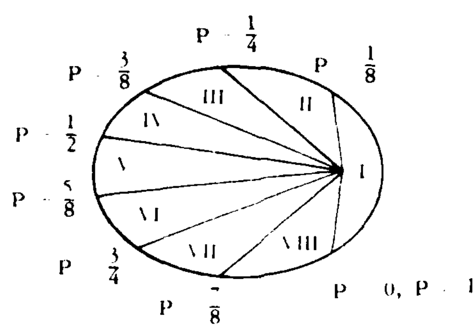


Figure 2-3. Law of Areas as applied to an elliptical orbit.

The squares of the sidereal periods* of any two planets are to each other as the cubes of their mean distances from the center of the sun.

To fit an earth orbital system, we should restate Kepler's third law as follows: *the squares of the periods of the orbits of two satellites are proportional to each other as the cubes of their mean distances from the center of the earth.* The mean distance is the length of the semimajor axis (a) of the ellipse, which is an average of the distances to perigee and apogee. In a circular orbit, the mean distance is the radius, r.

Newton's Laws

While Kepler was working out his three laws of planetary motion, Galileo, an Italian physicist and astronomer, was studying the effects of gravity on falling bodies. Newton drew upon the work of both Kepler and Galileo to formulate his laws of motion.

Newton's first law states:

Every body continues in a state of rest or of uniform motion in a straight line, unless it is compelled to change that state by a force imposed upon it.

In other words, a body at rest tends to remain at rest, and a body in motion tends to remain in motion unless an outside force acts upon it. We refer to this law sometimes as the law of inertia.

The second law of motion as stated by Newton says:

When a force is applied to a body, the time rate of change of momentum is proportional to, and in the direction of, the applied force.

If the mass remains constant, we can write this law as $F = Ma$.

Newton's third law of motion is the law of action and reaction:

For every action there is a reaction that is equal in magnitude but opposite in direction to the action.

If body A exerts a force on body B, then body B exerts an equal force in the opposite direction on body A.

Force as Measured in the English System

Newton stated the three laws of motion in terms of four quantities: force, mass, length, and time. Three of these, length, time, and either force or mass, may be completely independent. We define the fourth in terms of the other three by Newton's second law. Since Newton did not know the units and relative values of these quantities, Newton stated his second law as a proportionality. Assuming that mass does not change, we can state this proportionality as $F \propto ma$. If we select the proper units, we can write this statement as an equation:

$$F = ma$$

We use the following in the metric system of measurement:

F (dynes) = m (grams) times a (centimeters per second per second)
F (newtons) = m (kilograms) times a (meters per second per second)

*The period of a planet about the sun

The most common force experienced is that of weight, the measure of the body's gravitational attraction to the earth or other spatial body. Since this attraction is toward the center of the earth, weight, like any force, is a vector quantity. When the only force concerned is weight, we normally call the resulting acceleration g, the acceleration due to gravity. For this special case, we can write Newton's second law as:

$$W = Mg$$

We can use this equation as a definition of mass. The value of *g* near the surface of the earth is approximately 32.2 feet per second per second; g is a vector quantity since it is directed always toward the center of the earth. If we express the weight, W, in pounds, rearranging gives:

$$M = \frac{W \text{ (pounds)}}{g \text{ (ft/sec}^2)}$$

A slug is the term used for the unit of mass in this equation. Note that mass is a scalar quantity* and is an inherent property of the amount of matter in a body. Mass is independent of the gravitational field, whereas weight is dependent upon the field, the position in the field, and the mass of the body being weighed.

Finally, we can write Newton's second law as:

$$F \text{ (pounds)} = M \text{ (slugs) times } a \text{ (ft/sec}^2)$$

The following example shows the use of this system of units and the magnitude of the slug:

A package on earth weighs 161 pounds.

Find:

1. its mass in slugs.
2. the force necessary to just lift it vertically from a surface.
3. the force necessary to accelerate it 10 ft/sec² on a smooth, level surface.
4. its weight if it were on the moon; assume the value of "g" there is 1/6 of that value here on the earth.
5. its mass on the moon.

Solution:

1. $M \text{ (slugs)} = \frac{W}{g} = \frac{161 \text{ pounds}}{32.2 \text{ ft/sec}^2} = 5 \text{ slugs}$
2. $F = W = M g = (5 \text{ slugs})(32.2 \text{ ft/sec}^2) = 161 \text{ pounds}$
   The force must be applied upwards, in the direction opposite to weight.
3. $F = M a = (5 \text{ slugs})(10 \text{ ft/sec}^2) = 50 \text{ pounds}$
4. $W_{moon} = M\, g_{moon} = (5 \text{ slugs})\frac{32.2}{6} \text{ ft/sec}^2) = 26.83 \text{ pounds}$
5. $M_{moon} = \frac{W_{moon}}{g_{moon}} = \frac{26.83 \text{ pounds}}{\frac{32.2}{6} \text{ ft/sec}^2} = 5 \text{ slugs}$

*A scalar quantity has magnitude *only*, in contrast to a vector quantity which has magnitude and direction

This last solution is to reemphasize that mass is independent of position. We will show later that the local value of g varies with altitude above the earth. It is significant to note that the weight will vary in such a way that the ratio $\frac{W}{g}$ remains constant.

Energy and Work

We can define work, w, as the product of the component of force in the direction of motion and the distance moved. Thus, if we applied a force, F, and a body moves a distance, s, in the direction that we applied the force, w = Fs (fig. 2-4). The units of work are foot-pounds. Work is a scalar as distinguished from a vector quantity.

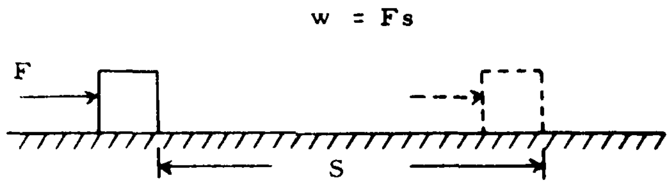


Figure 2-4. Work performed as a force (F) is moved over the distance s.

To do work against gravity, we must apply a force to overcome the weight, which is the force caused by gravitational acceleration, g. Therefore, F = Mg. If we lift the body a height, h, (fig. 2-5), and friction is negligible, w = Mgh. For problems in which h is much less than the radius from the center of the earth ($h \ll r$), we can consider g a constant.

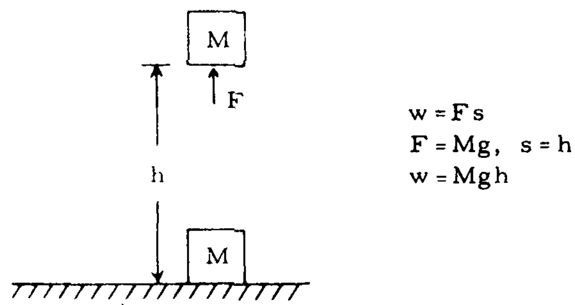


Figure 2-5. Work performed in lifting.

If we push an object up a frictionless inclined plane, the work done is still Mgh (fig. 2-6).

For orbital mechanics problems, g varies and the value of $\sqrt{g_1 g_2}$ must replace g where the subscripts indicate the beginning and final values of g. In such cases

$$w = M\sqrt{g_1 g_2}\,h$$

$$w = Fs$$
$$F = Mg \sin \theta$$
$$s = \frac{h}{\sin \theta}$$
$$w = Mgh$$

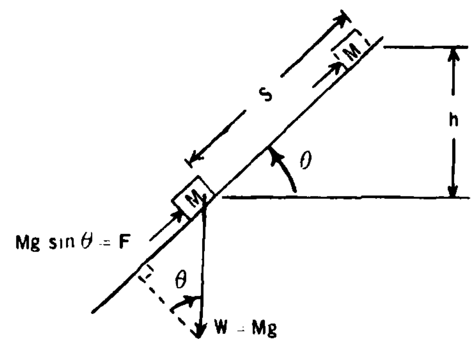


Figure 2-6. Work performed on a frictionless inclined plane.

Another type of work is that work done against inertia. If, in moving from one point to another, we change the velocity of a body, work is done. We compute this work against inertia by using the following steps:

$$w = Fs$$
$$\text{but, } F = Ma$$
$$\text{and } 2as = v_f^2 - v_o^2$$
$$s = \frac{v_f^2 - v_o^2}{2a}$$
$$\text{so, } w = Fs = M\,\frac{a\,(v_f^2 - v_o^2)}{2a} = \frac{M\,(v_f^2 - v_o^2)}{2} = \frac{Mv_f^2}{2} - \frac{Mv_o^2}{2}$$

We define the quantity $\frac{Mv^2}{2}$ as kinetic energy (KE). Therefore, work done against inertia (if the altitude and the mass remain the same) is equal to the change in kinetic energy. We define energy as the ability to do work, and it is obvious that a moving body has the ability to do work (for example, a moving hammer's ability to drive a nail). A body is also able to do work because of its position or altitude. We know this as potential energy (PE). Units used to measure energy are similar to those used to measure work in that both are scalar rather than vector quantities. The sum of the kinetic and potential energy of a body is its total mechanical energy.

Newton's Law of Universal Gravitation

Newton published his *Principia* in 1687 and included in it the law of universal gravitation, which he had been considering for approximately 20 years. Newton based this law on his own observations. Later work showed that it was only an approximation, but an extremely good approximation. The law states:

Every particle in the universe attracts every other particle with a force that is proportional to the product of the masses and inversely proportional to the square of the distance between the particles.

Newton introduced a constant of proportionality, G, termed the Universal Gravitational Constant, and wrote the law in this manner:

$$F = \frac{Gm_1m_2}{r^2}$$

Cavendish first determined the value of G, the Universal Gravitational Constant, in a classical experiment using a torsion balance. The value of G is quite small ($G = 6.6695 \times 10^{-8}$ cgs units). In most problems the mass of one of the bodies is quite large. It is convenient to combine G and the large mass, m_1, into a new constant, μ (mu), which we define as the gravitational parameter. This parameter has different values depending upon the value of the large mass, m_1. If m_1 refers to the earth, the gravitational parameter, μ, will apply to all earth satellite problems. However, if the problem concerns satellites of the sum or other large bodies, μ will have a different value based on the mass of that body.

If we simplify the law of gravitational attraction by combining G and m_1 and by adjusting the results for the English engineering unit system, we obtain the following:

$$G\,m_1 = \mu \frac{ft^3}{sec^2}$$

$$F = \frac{\mu}{r^2} m \text{ (Where F is lb force and m is slugs)}$$

If this expression is equated to the expression of Newton's second law of motion, as it applies in a gravitational field, we see that:

$$F = mg = \frac{\mu}{r^2} m$$

and after dividing by the unit mass, m, we obtain:

$$g = \frac{\mu}{r^2}$$

Thus, the value of g varies inversely as the square of the distance from the center of the attracting body.

For problems involving earth satellites, the following two constants are necessary for a proper solution:

$$G\,m_{earth} = \mu_{earth} = 14.08 \times 10^{15} \frac{ft^3}{sec^2}$$

$$r_e \text{ (radius of earth)} = 20.9 \times 10^6 \text{ ft}$$

We must use the formulas with proper concern for the units involved, and the value given for μ applies only to bodies attracted to the earth.

Before applying Newton's law of universal gravitation to the solution of problems, we should consider the possible paths that a body in unpowered flight must follow through space.

CONIC SECTIONS

The Greek mathematicians studied the conic sections, and, since then, scientists have accumulated a body of knowledge concerning them. They have assumed new significance in the field of astronautics because a conic section can represent any free-flight trajectory. The study of conic sections, or conics, is part of analytic geometry, a branch of mathematics that brings together concepts from algebra, geometry, and trigonometry.

A conic section is a curve formed when a plane cuts through a right circular cone at any point except at the vertex, or center. If the plane cuts both sides of one nappe of the cone, the section is an *ellipse* (fig 2-7). The *circle* is a special case of the ellipse occurring when the plane cuts the cone perpendicularly to the axis. If the plane cuts the cone in such a way that it is parallel to one of the sides of the cone, the section is a *parabola*. If the plane cuts both nappes of the cone, the section is a *hyperbola* that has two branches.

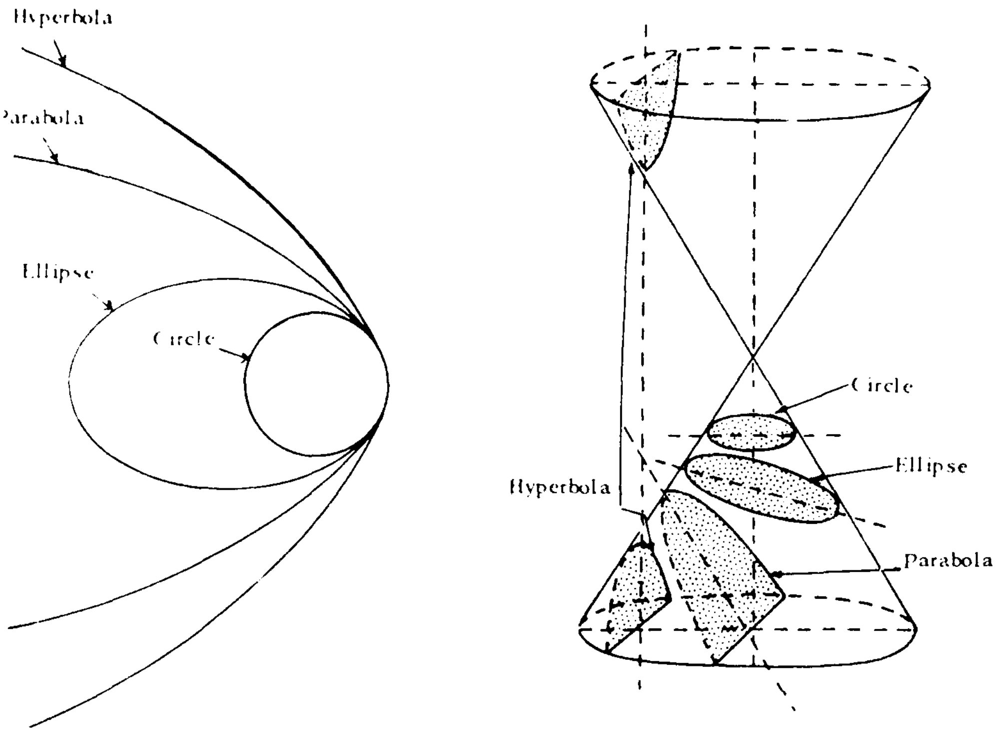


Figure 2-7. Conic sections.

In one mathematical sense, we can define all conic sections in terms of eccentricity (ϵ). The numerical value of ϵ is an indication of the relative shape of the conic (rotund or slender) and an indication of the identity of the conic. If the eccentricity is zero, the conic is a circle; if the eccentricity is greater than zero but less than one, the conic is an ellipse; if the eccentricity is equal to one, the conic is a prabola; and if the eccentricity is greater than one, the conic is a hyperbola.

Conic Sections and the Coordinate Systems

In locating orbits or trajectories in space, it is possible to use either rectangular (sometimes) called Cartesian) or polar coordinates. In dealing with artificial satellites, it is often more convenient to use polar rather than rectangular coordinates because both the origin of the coordinates and one of the foci of the ellipse can be the center of the earth.

If we superimpose rectangular and polar coordinates upon a set of conics as shown in figure 2-8, we can derive equations of the curves. The formula for the eccentricity of a conic is $\epsilon = \frac{r}{d}$. This ratio is constant for a specific curve.

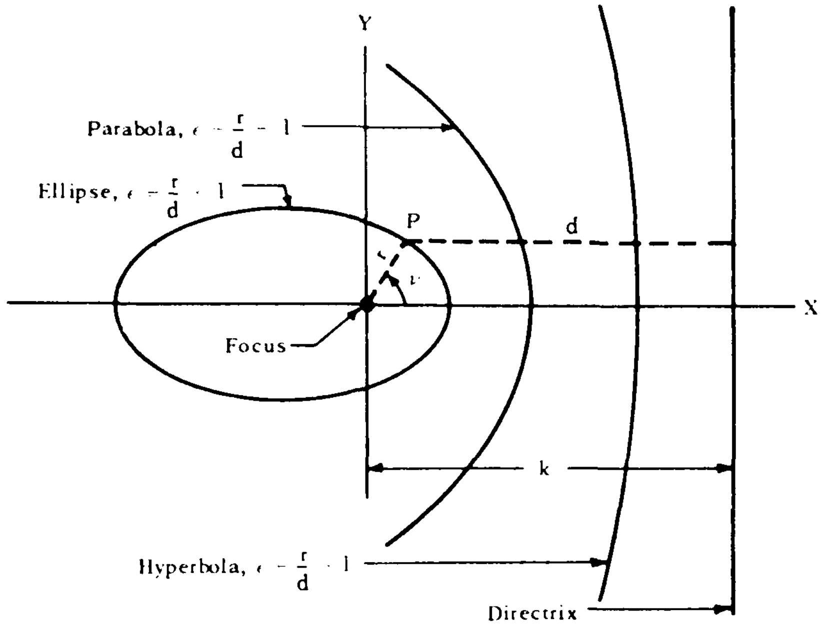


Figure 2-8. Rectangular and polar coordinates superimposed on the conic sections.

In Cartesian coordinates:

$$\epsilon = \frac{r}{d} = \frac{\sqrt{x^2 + y^2}}{k - x}$$

$$\sqrt{x^2 + y^2} = \epsilon(k - x)$$

Squaring both sides gives:

$$x^2 + y^2 = \epsilon^2(k - x)^2$$

To convert the rectangular coordinates to polar coordinates, substitute as follows:

$$x = r\cos\nu \quad (\nu \text{ is the lower case Greek nu})$$

$$y = r\sin\nu$$

$$\epsilon = \frac{r}{d} = \frac{r}{k - r\cos\nu}$$

$$k\epsilon - r\epsilon\cos\nu = r$$

$$r + r\epsilon\cos\nu = k\epsilon$$

$$r = \frac{k\epsilon}{1 + \epsilon\cos\nu}$$

This result is the general equation for *all* conics.

Ellipse

The ellipse is the curve traced by a point (P) moving in a plane such that the sum of its distances from two fixed points (foci) is a constant. In the ellipse in figure 2-9, the following are shown: the foci (F and F'); c, distance from origin to either focus; a, distance from origin to either vertex (semimajor axis); 2a, major axis; b, distance from origin to intercept on y-axis (semiminor axis); 2b, minor axis; and r + r', distances from any point (P) on the ellipse to the respective foci (F and F').

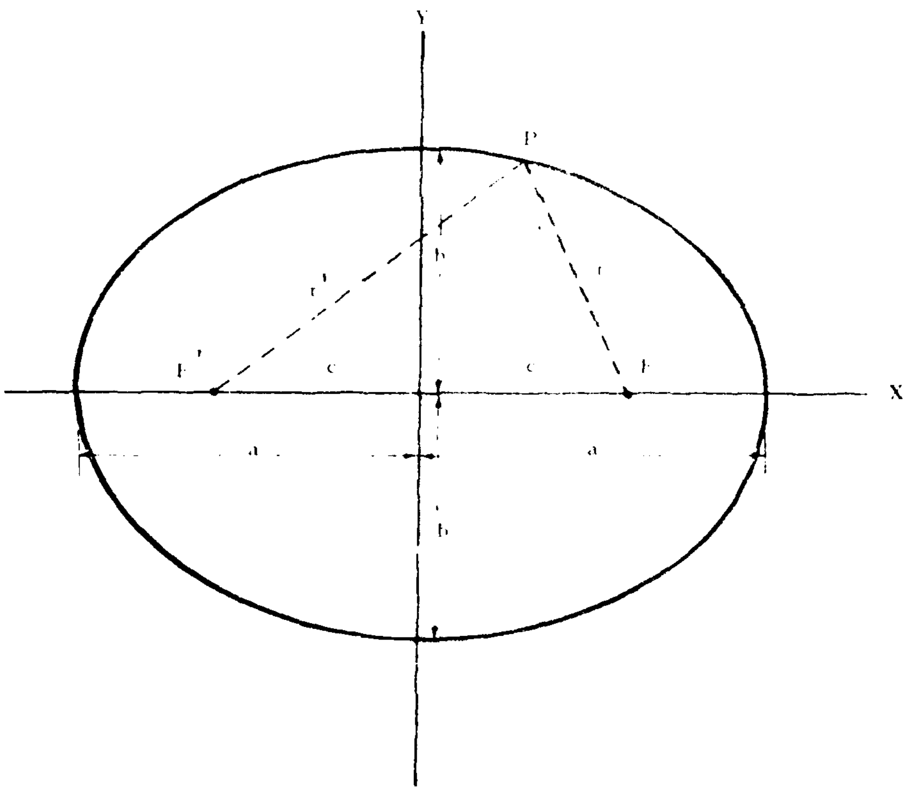


Figure 2-9. Ellipse with center at origin of rectangular coordinate system.

We can derive a number of relationships that are very useful in astronautics from the geometry of the ellipse:

$$r + r' = 2a \text{ (at any point on the ellipse)}$$

$$a^2 = b^2 + c^2 \text{ or } a = \sqrt{b^2 + c^2}$$

$$b = \sqrt{a^2 - c^2}$$

$$c = \sqrt{a^2 - b^2}$$

The eccentricity of the ellipse $(\epsilon) = \frac{c}{a}$. A chord through either focus perpendicular to the major axis is the *latus rectum* and its length $= \frac{2b^2}{a}$.

We can use these relationships to determine the parameters of an elliptical orbit of a satellite when we know only the radius of perigee and the radius of apogee. These parameters are important because, as we show later, they are related to the total mechanical energy and total angular momentum of the satellite. They offer a means of determining these values through the simple arithmetic of an ellipse rather than the vector calculus of celestial mechanics.

Sample problem: A satellite in a transfer orbit has a perigee at 300 nautical miles (NM) above the surface of the earth and an apogee at 19,360 NM. Find a, b, c, and ϵ for the ellipse traced out by this satellite.

Solution:

Since the center of the earth is one focus of the ellipse, first convert the apogee and perigee to radii by adding the radius of the earth (3,440 NM):

$$\begin{aligned} \text{radius of perigee } r_p &= \text{altitude of perigee} + \text{radius of earth} \\ &= 300 + 3{,}400 = 3{,}740 \text{ NM.} \\ \text{radius of apogee } r_a &= \text{altitude of apogee} + \text{radius of earth} \\ &= 19{,}360 + 3{,}440 = 22{,}800 \text{ NM.} \end{aligned}$$

With this information, we can make an exaggerated sketch of the ellipse (fig. 2-10). Compare this with figure 2-9 to obtain:

$$r_p + r_a = \text{major axis} = 2a$$

then $\quad 2a = 3{,}740 + 22{,}800 = 26{,}540$ NM

or $\quad a = \dfrac{26{,}540}{2} = 13{,}270$ NM

Also from comparing figures 2-9 and 2-10:

$$\begin{aligned} c &= a - r_p \\ &= 13{,}270 - 3{,}740 = 9{,}530 \text{ NM} \end{aligned}$$

Since a and c are known, find b from the relationship given:

$$b = \sqrt{a^2 - c^2}$$

$$\text{or } b = \sqrt{(1.327 \times 10^4)^2 - (.953 \times 10^4)^2}$$

$$b = \sqrt{(1.761 \times 10^8) - (.908 \times 10^8)} = \sqrt{.853 \times 10^8}$$

$$b = .923 \times 10^4 = 9{,}230 \text{ NM}$$

According to the formula given for eccentricity:

$$\epsilon = \frac{c}{a}$$

$$\epsilon = \frac{9{,}530}{13{,}270} = .718$$

The ellipse is a conic section with eccentricity less than 1 ($\epsilon < 1$).

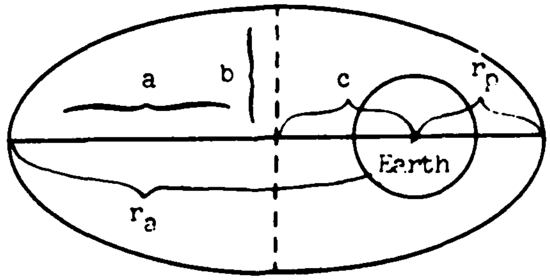


Figure 2-10. Orbit of an artificial satellite showing radius of perigee and radius of apogee (not to scale).

The circle is a special case of an ellipse in which the foci have merged at the center; thus $\epsilon = 0$. We can use the ellipse relationships for a circle.

ENERGY AND MOMENTUM

Once we understand the basic geometry of a trajectory or orbit, the next subject for investigation is the physics of energy and momentum. From concepts of linear and angular motion, concepts of linear and angular momenta logically follow. Once we delineate the formulas for computing the specific angular momentum and the specific mechanical energy of a body in orbit then we can solve for unknown quantities, such as the altitude of the body above the surface of the earth or the velocity at any point on the orbit. The laws of conservation of specific mechanical energy and specific angular momentum govern any body in space following a free-flight path, whether it is a missile, a satellite, or a natural body. Once we know the value of either of these items at any point along a free-flight trajectory or orbit, then we know its value at all other points, since the value does not change unless some outside force acts upon the body.

Mechanical Energy

The law of conservation of energy states that *energy can neither be created nor destroyed but only converted from one form to another.* We can apply this law to orbital mechanics and restate it this way:

The total mechanical energy of an object in free motion is constant, provided that no external work is done on or by the system.

During reentry, the system does some work and converts some of the mechanical energy to heat. Similarly, during the launch, work is done on the system as the propulsion units give up chemical energy. In this chapter, we consider only the free-flight portion of the trajectory, and we assume that there is no thrust and no drag.

To establish a common understanding about changes in the amount of energy, it is necessary to agree upon a zero reference point for energy. We can, and often do, measure potential energy, or energy due to position, from sea level. In working with earth-orbiting systems, however, the convention is to consider a body as having zero potential energy if it is at an infinite distance from the earth and as having zero kinetic energy if it is absolutely at rest with respect to the center of the earth. Under these circumstances, the total mechanical energy (PE + KE) is equal to zero. If the total mechanical energy is positive—that is, larger than zero—the body has enough energy to escape from the earth. If the total mechanical energy is negative—that is, less than zero—the body does not have enough energy to escape from the earth, and it must be either in orbit or on a ballistic trajectory.

The formula for PE, with the reference system as stated above, is $PE = -\frac{m\mu}{r}$. Instead of using potential energy (PE), we can use a specific PE (PE per unit mass) if we divide both sides by m; for example

$$\frac{PE}{m} = -\frac{\mu}{r}$$

If a body is at infinity, it has a specific PE equal to $-\frac{\mu}{r} = \frac{\mu}{\infty} = 0$.

We can present a similar case for kinetic energy (KE). A body with some velocity relative to the center of the earth has kinetic energy defined by:

$$KE = \frac{mv^2}{2}$$

Again, we can define the specific kinetic energy (kinetic energy per unit mass) as:

$$\text{Specific } KE = \frac{KE}{m} = \frac{v^2}{2}$$

In general, a body in free motion in space has a particular amount of mechanical energy, and this amount is constant because of the conservation of mechanical energy.

$$\text{Total Mechanical Energy} = KE + PE$$

We can obtain a more useful expression if we define specific mechanical energy, E, or the total mechanical energy per unit mass. Thus we can write:

$$E = \frac{\text{Total Mechanical Energy}}{m}$$

$$E = \frac{KE}{m} + \frac{PE}{m}$$

$$E = \frac{v^2}{2} - \frac{\mu}{r}$$

Specific mechanical energy, E, is conserved in unpowered flight in space. The units of E are $\frac{ft^2}{sec^2}$.

Since the mass term does not appear directly in the equation, E represents the specific mechanical energy of a body in general.

If the solution to the specific mechanical energy equation yields a negative value for E, the body is on an elliptical or circular path (nonescape path). If E exactly equals zero, the path is parabolic. This is the minimum energy escape path. If E is positive, the path is hyperbolic, and the body will escape from the earth's gravitational field.

Although the value of E, once determined, remains constant in free flight, there is a continuous change in the values of specific KE and specific PE. High velocities nearer the surface of the earth, representing high specific KE, are exchanged for greater specific PE as distance from the center of the earth increases. In general, the spacecraft trades velocity for altitude; KE for PE.

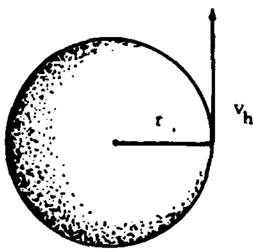
Linear and Angular Momentum

When a body is in motion, it has momentum. Momentum is the property a body possesses because of its mass and its velocity. In linear motion, momentum is expressed as mv and has the units. $\frac{\text{foot-slug}}{\text{sec}}$.

When a rigid body, such as a flywheel, rotates about a center, it has angular momentum. Once the flywheel is in motion, its angular momentum remains constant if forces such as friction and air resistance do not act upon it. Similarly, a gyroscope will rotate indefinitely in the absence of friction and air resistance. Thus, ignoring such losses, angular momentum will remain constant. In space, the assumption is that such forces are negligible and that angular momentum is conserved. This is another tool to use in analyzing orbital systems.

Angular momentum is the product of moment of inertia, I, and the angular velocity, ω. The formula, $m\ r^2$, expresses the momentum of inertia of a body of mass, m, rotating about a center at a distance, r. The angular momentum equals $m\ r^2\ \omega$.

For convenience in calculations, the term, specific angular momentum, H, is defined as the angular momentum per unit mass. Remembering that the magnitude of the instantaneous velocity vector of a body rotating in constant circular motion about a center with radius r equals ωr and that the vector is perpendicular to the radius, we can simplify the expression for specific angular momentum of a circular orbit as shown in figure 2-11.



$$\text{Angular Momentum} = m\ r^2\ \omega$$

$$\omega = \frac{v_{horizontal}}{r}$$

$$H = \frac{\text{Ang. Momentum}}{m}$$

$$\therefore H = v_h r \text{ (Circular Motion)}$$

Figure 2-11. Specific angular momentum of a circular orbit.

The general application of specific angular momentum to all orbits requires that we use the component of velocity perpendicular to the radius vector. We define this velocity component as

$$v_h = v \cos \phi$$

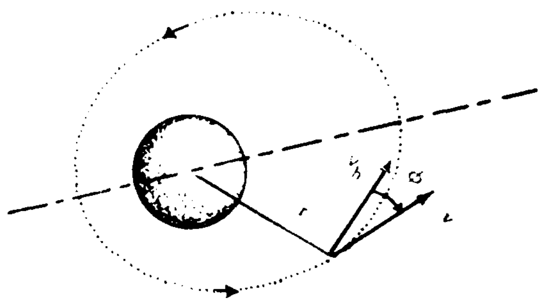
where ϕ is the angle the velocity vector makes with the local horizontal, a line perpendicular to the radius. In an elliptical orbit, the geometry is as shown in figure 2-12. The body in orbit has a total velocity, v, that is always tangent to the flight path.

The formula, $H = v\ r \cos \phi$, defines the specific angular momentum for all orbital cases. The angle ϕ *is the flight path angle* and is the angle between the local horizontal and the total velocity vector. We should note that the angle, ϕ, equals zero for circular orbits. In elliptical orbits, ϕ is zero at the points of apogee and perigee.

The two important formulas presented in this section are those for E and H. These formulas permit a complete definition of a trajectory or an orbit from certain basic data:

$$E = \frac{v^2}{2} - \frac{\mu}{r}, \text{ when units of E are } \frac{ft^2}{sec^2}$$

$$H = v\ r \cos \phi, \text{ when units of H are } \frac{ft^2}{sec}$$



$$H = v_h r$$

$$\text{but } v_h = v \cos\phi$$

$$\therefore H = v r \cos\phi \text{ (General orbit)}$$

Figure 2-12. Specific angular momentum.

If we know v, r, and ϕ for a given trajectory (or orbit) at a given position, then we can determine E and H. In the absence of outside forces, E and H are constants; therefore we can determine v, r, and ϕ at any other position on the trajectory or orbit. We can use equations for the specific angular momentum and the specific mechanical energy in practical application to the two-body problem and to the free-flight portion of the ballistic missile trajectory.

THE TWO BODY PROBLEM

It is implicit in Newton's law of universal gravitation that every mass unit in the universe attracts, and is attracted by, every other mass unit in the universe. Clearly, small masses at large distances infinitesimally attract each other. It is neither feasible nor necessary to consider mutual attraction of a large number of bodies in many astronautics problems. The most frequent problems of astronautics involve only two interacting bodies: a missile payload or satellite, and the earth. In these instances, the sun and moon effects are negligible except in the case of a space probe. The moon will noticeably affect a space probe as the probe passes close to the moon. The sun will control the probe if it escapes from the earth's gravitational field.

The primary interests of military officers concerned with operational matters are launching a missile from one point on the earth's surface to strike another point on the earth's surface and launching earth satellites. In these problems, the path followed by the payload is described adequately by considering only two bodies, the earth and the payload. The problem of two bodies is termed the two-body problem; its solution dates back to Newton.

It is fortunate that the solution of the two-body trajectory is simple and straightforward. A general solution to a trajectory involving more than two bodies does not exist. Special solutions for these more complex trajectories usually require machine calculation.

Figure 2-13 graphically describes the two-body problem. A small body m_2, has a velocity, v, at a distance, r, from the origin chosen as the center of mass of a very massive body, m_1. The problem is to

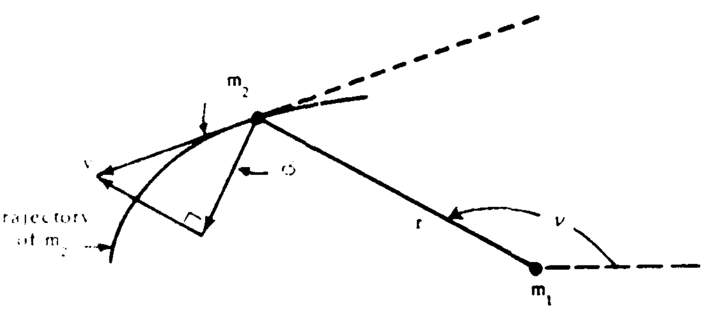


Figure 2-13. Trajectory relationships.

establish the path followed by body, m_2, or to define its trajectory. This is a typical problem in mechanics: given the present conditions of a body, what will these conditions be at any time, t, later? First, we shall find r as a function of ν, where ν is the polar angle measured from a reference axis to the radius vector.

At the outset, it should be apparent that the entire trajectory will take place in the plane defined by the velocity vector and the point origin. There are no forces causing the body, m_2, to move out of this plane; otherwise the conditions are not those of a two-body free-flight problem.

In the earlier outline of the laws of conservation of energy and momentum, we established the following conditions:

$$\frac{v^2}{2} - \frac{\mu}{r} = E = \text{a constant} \tag{1}$$

$$H = vr\cos\phi = \text{a constant} \tag{2}$$

We can combine equations 1 and 2 and, with the aid of calculus, we can derive the following equation:

$$r = \frac{H^2/\mu}{1 + \sqrt{1 + \frac{2EH^2}{\mu^2}}\cos\nu} \tag{3}$$

Equation 3 is the equation of a two-body trajectory in polar coordinates.

Earlier we gave the following equation as the equation of any conic section in polar coordinates with the origin located at a focus:

$$r = \frac{k\epsilon}{1 + \epsilon\cos\nu} \tag{4}$$

Equations 3 and 4 are of the same form; hence, equation 3 is the equation of any conic section (origin at a focus) in terms of the physical constants, E and H, and the two-body trajectories are then conic sections. This conclusion substantiates Kepler's first law. In fact, Kepler's first law is a special case because an ellipse is just one form of conic section.

Since equations 3 and 4 are of the same form, it is possible to equate like terms, which will lead to relationships between the physical constants, E, H, and μ, and the geometrical constants, ϵ, a, b, and c. Thus:

$$k\epsilon = \frac{H^2}{\mu} \tag{5}$$

and

$$\epsilon = \sqrt{1 + \frac{2EH^2}{\mu^2}} \tag{6}$$

Physical Interpretation of the Two-Body Trajectory Equation

If we analyze the two-body trajectory, we will get an understanding of the physical reaction of a vehicle (small body) under the influence of a planet (large body).

If $E < 0$, the trajectory is an ellipse. What is the condition that E be less than zero? It is simply that the kinetic energy of the small mass, m_2, because of its relatively low velocity, is less than the

magnitude of its potential energy. Therefore, the body cannot possibly go all the way to infinity; that is, it cannot go to a point where a larger body no longer attracts it—where the potential energy is zero. The smaller mass cannot escape. It must remain "captured" by the force field of the larger body. Therefore, the force field of the larger body will turn the smaller mass toward the larger body. In keeping with the idea of potential energy, the smaller body will always fall toward the more massive body. When this particular balance of energy exists, the trajectory is elliptical with one focus coincident with the center of mass of the larger body. In the actual physical case, the larger body will have a finite size: that is, it will not be a point mass, and this ellipse may intersect the surface of the larger body as it does in the case of a ballistic missile. If the velocity is sufficiently high, and its direction proper, the ellipse may completely encircle the central body, the condition of the satellite.

If $E = 0$, the kinetic energy exactly equals the magnitude of the potential energy, and the small mass, m_2, has just enough energy to travel to infinity, away from the influence of the central body, and come to rest there. The small body will follow a parabolic path to infinity. The velocity that we associate with this very special energy level is also very special. It is commonly called the "escape velocity."

We can calculate escape velocity by setting $E = 0$ in the mechanical energy equation 1 as follows:

$$\frac{v_{esc}^2}{2} - \frac{\mu}{r} = E = 0$$

$$v_{esc}^2 = \left(2\frac{\mu}{r}\right)$$

$$v_{esc} = \sqrt{\frac{2\mu}{r}} \tag{7}$$

Thus, we can see that escape velocity decreases with distance from the center of the earth. At the earth's surface,

$$v_{esc} = \sqrt{\frac{2\mu}{r_e}} = \left[\frac{(2)(14.08)(10^{15}\frac{ft^3}{sec^2})}{(20.9)(10^6\ ft)}\right]^{1/2}$$

$$v_{esc} = 36,700\ ft/sec.$$

If the velocity of the small mass exceeds escape velocity, which will be the case if $E > O$, it will follow a hyperbolic trajectory to infinity. In practice, infinity is a large distance at which the earth's attractive force is insignificant, and there the mass will have some residual velocity. In a mathematical sense, the body would continue to have velocity at infinity. In a physical sense, it would have velocity relative to the earth at any large distance from the earth.

Considering the sounding rocket, only the straight-line, degenerate conic is a possible trajectory. But, again, the value of E will determine whether escape is possible; if $E < O$, the straight-line trajectory cannot extend to infinity. If $E = O$ or $E > O$, the straight line will extend to infinity.

Example Problem: The first US "moon shot," the Pioneer I, attained a height of approximately 61,410 nautical miles above the earth's surface. Assuming that the Pioneer had been a sounding rocket (a rocket fired vertically), and assuming a spherical, nonrotating earth without atmosphere, calculate the following:

1. E (total specific energy)
2. Impact velocity (earth's surface)

Solution: Given

1. At apogee (greatest distance from earth):

   Altitude (above earth's surface) = 61,410 NM

   Earth radius = 3,440 NM

   Velocity = 0 (Only for a sounding rocket)

   r = altitude + earth's radius

   r = (61,410 + 3,440 NM = 64,850 NM

$$E = \frac{v^2}{2} - \frac{\mu}{r} = 0 - \frac{(14.08)(10^{15})\,\frac{ft^3}{sec^2}}{(64{,}850)\ NM\ (6{,}080)\,\frac{ft}{NM}}$$

$$E = -3.57 \times 10^7\ ft^2/sec^2$$

Answer

2. Since the specific energy is constant,

   At the earth's surface:

$$r = 3{,}440\ NM$$

$$E = -3.57 \times 10^7\ ft^2/sec^2$$

$$E = \frac{v^2}{2} - \frac{\mu}{r}$$

$$-3.57 \times 10^7\,\frac{ft^2}{sec^2} = \frac{v^2}{2} - \frac{(14.08)(10^{15}\,\frac{ft^3}{sec^2})}{(3{,}440)\ NM\ (6{,}080)\,\frac{ft}{NM}}$$

$$v^2_{(impact)} = 2[(-3.57 \times 10^7) + (67.4 \times 10^7)] = 2(63.8 \times 10^7)\,ft^2/sec^2$$

$$v_{(impact)} = 35{,}700\ ft/sec$$

Answer

This is the approximate burnout velocity of the vehicle. As the surface escape velocity is 36,700 ft/sec, it is clear that Pioneer I did not attain escape velocity, and so it returned to earth.

Elliptical Trajectory Parameters

While parabolic and hyperbolic trajectories, especially the latter, are of interest in problems of interplanetary travel, elliptical trajectories comprise the ballistic missile and satellite cases, which are of current military interest. It is important to relate the dimensions of an ellipse (a, b, and c) to the physical constants (E, H, and μ) as we previously did for ϵ.

We presented the relationship, $r_a + r_p = 2a$, earlier. If we applied this equation to point P in figure 2-14,

$$r = \frac{k\epsilon}{1 + \epsilon \cos \nu}\ (\nu = 0 \text{ at P})$$

$r = \frac{k\epsilon}{1 + \epsilon}$ and $r_p = a - c$. Then,

$$\frac{k\epsilon}{1 + \epsilon} = a - c \qquad (8)$$

But $\epsilon = \frac{c}{a}$; therefore, substituting $c = \epsilon a$ into 8,

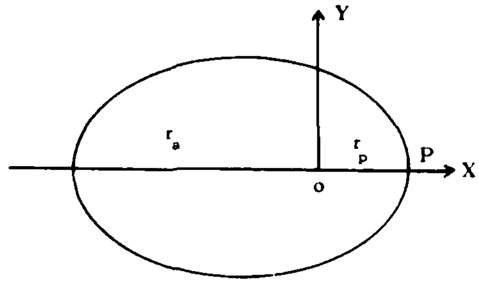


Figure 2-14. Ellipse.

$$\frac{k\epsilon}{1+\epsilon} = a - \epsilon a = a(1-\epsilon)$$

$$k\epsilon = a(1-\epsilon^2) \tag{9}$$

But from equations 5 and 6,

$$k\epsilon = H^2/\mu \text{ and } \epsilon = \sqrt{1 + \frac{2EH^2}{\mu^2}}$$

Substituting these relationships into equation 9,

$$\frac{H^2}{\mu} = a \quad 1 - 1 - \frac{2EH^2}{\mu^2}$$

$$\frac{H^2}{\mu} = -\frac{2EH^2 a}{\mu^2}$$

$$-1 = \frac{2Ea}{\mu}$$

$$a = -\frac{\mu}{2E} \text{(extremely useful)} \tag{10}$$

$$\text{Also } a = \frac{\mu}{\left(2\ \frac{v^2}{2} - \frac{\mu}{r}\right)} \tag{11}$$

From the following equation:

$$\frac{b^2}{a^2} = 1 - \epsilon^2$$

$$\frac{b^2}{a} = a(1-\epsilon^2)$$

But from equation 9,

$$a(1-\epsilon^2) = k\epsilon = \frac{H^2}{\mu}$$

Therefore,

$$\frac{b^2}{a} = \frac{H^2}{\mu} \tag{12}$$

Equations 10 and 12 are extremely important relationships. An understanding of them is essential to material that follows on ballistic missiles and satellites. If injection conditions of speed and radius are fixed, it is clear that a, the semimajor axis of the elliptical trajectory, becomes fixed, regardless of the value of the flight path angle at burnout. Equation 10 points out there is a direct relationship between the size of an orbit and the energy level of the orbiting object. Equation 12 points out that for a given energy level there is a direct relationship between the length of the semi latus recturm of an elliptical trajectory (a shape parameter) and the specific angular momentum of the orbiting object. This implies that E and H determine the size and shape of an elliptical trajectory.

Two-Body Trajectory Definitions and Geometry

We have introduced the general equation of two-body trajectories. Before proceeding to problem applications we should consider in detail some commonly used terms and symbols.

First, refer to figure 2-15. In general, the periapsis is the point P and the apoapsis is P′. If the earth is at point O, the ellipse represents the trajectory of an earth satellite; then the perigee becomes P and the apogee becomes P′. If the sun is at point O, the ellipse would represent a planetary orbit, then the perihelion is P and the aphelion is P′.

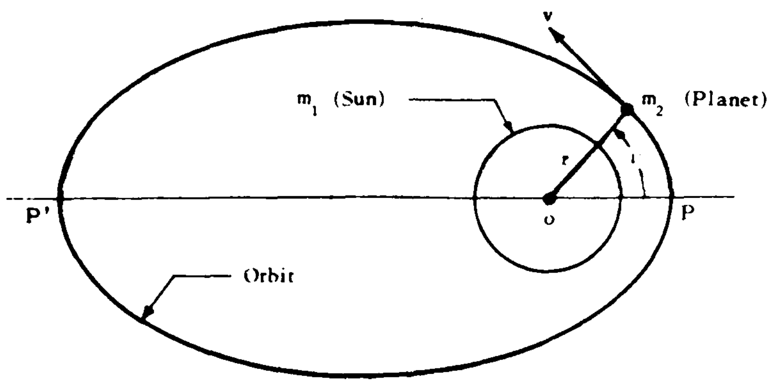


Figure 2-15. Sun-centered orbit.

To explain the use of the angle ν, we will discuss briefly the geometry of satellite orbits. Figure 2-15 depicts a planetary orbit (not to scale). In astronomy and celestial mechanics, standard practice is to measure a body's position from perihelion point P. There are several reasons for using perihelion, including the fact that perihelion of any body in the Solar System except Mercury and Venus is closer to the Earth's orbit than is the body's aphelion. In fact for a highly eccentric orbit such as a comet's, we could not see the body at aphelion. To conform to accepted practice, we have introduced the angle nu (ν), measured from periapsis. This angle, called the true anomaly, is of considerable importance in time-of-flight calculations. The true anomaly does not lend itself well to ballistic missile problems, as we can see from the ballistic missile geometry in figure 2-16.

In a ballistic missile trajectory, perigee is entirely fictitious. The missile obviously never traverses the dashed portion of the trajectory. The solid portion of the trajectory is all that is of real interest,

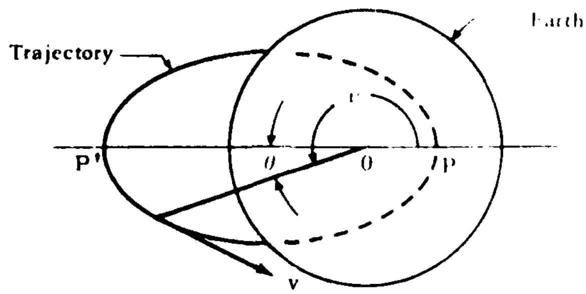


Figure 2-16. Ballistic missile trajectory.

and this portion is in the second and third quadrants of the angle ν. Then it is convenient to define an angle θ, measured counterclockwise from apogee (apoapsis, in general), such that

$$\nu = \theta + \pi. \tag{13}$$

Normally, θ will have values in the first and fourth quadrants. From 13, it is clear that we can interchange derivatives of ν and θ. By substituting 13 into 4, we can find the equation of a conic section in terms of θ,

$$r = \frac{k\epsilon}{1 + \epsilon \cos \nu} = \frac{k\epsilon}{1 + \epsilon \cos (\theta + \pi)}$$

$$r = \frac{k\epsilon}{1 - \epsilon \cos \theta}.$$

With this understanding of the relationship between ν and θ, we can use ν conveniently when working with satellite and space trajectories and θ when working with ballistic missiles.

EARTH SATELLITES

During their free-flight trajectory, satellites and ballistic missiles follow paths described by the two-body equation. For a satellite to achieve orbit, we must add enough energy to the vehicle so that the ellipse does not intersect the surface of the earth. However, we must not add enough energy to allow the vehicle to escape. Therefore, the ellipse and the circle are the paths of primary interest.

The orientation, shape, and size of orbits are important to the accomplishment of prescribed missions. Therefore, eccentricity (ϵ), major axis (2a), minor axis (2b), and distance between the foci (2c) are of interest. It is necessary to know the relationships of these geometric values to the orbital parameters to make an analysis of an orbit. For example, it is helpful to remember that:

$$r_p \text{ (radius at perigee)} = a - c$$

$$r_a \text{ (radius at apogee)} = a + c$$

$$r_p + r_a = 2a$$

$$\epsilon = \frac{c}{a}$$

$$a^2 = b^2 + c^2$$

Specific mechanical energy, E, and specific angular momentum, H, are of primary concern when we discuss elliptical and circular orbits. If there are no outside forces acting on a vehicle in an orbit, the specific mechanical energy and the specific angular momentum will have constant values, regardless of position in the orbit.

This means that if we know E and H at one point in orbit, we know them at each and every other point in orbit. If we know radius r, speed v, and flight angle ϕ at a given position, we can determine E and H from:

$$E = \frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}$$

$$H = vr \cos \phi$$

If we know the values of E and H for a particular orbit, and we are to determine the speed and flight path angle at a certain point in the orbit, we can solve the energy equation for v, and then, the angular momentum equation for ϕ.

The equations for the speed in circular and elliptical orbits are important. The equation for circular speed is:

$$v = \sqrt{\frac{\mu}{r}}$$

The equation for elliptical speed is:

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}$$

Another equation that is important in the analysis of orbits is the equation for orbital period. For a circular orbit the distance around is the circumference of the circle, which is $2\pi r$. Therefore, the period, which equals the distance around divided by the speed, is:

$$P = \frac{2\pi r}{v}.$$

Now, substitute for v the speed in circular orbit:

$$v = \sqrt{\frac{\mu}{r}}$$

$$P = \frac{2\pi r}{\sqrt{\frac{\mu}{r}}}$$

Multiply numerator and denominator of the right hand side by $\sqrt{r}$:

$$P = \frac{2\pi r \sqrt{r}}{\sqrt{\frac{\mu}{r}}\sqrt{r}} = \frac{2\pi r^{3/2}}{\sqrt{\mu}}$$

Using the principle of Kepler's third law, replace r by the mean distance from the focus, which equals the semimajor axis a, and the equation becomes

$$P = \frac{2\pi a^{3/2}}{\sqrt{\mu}}$$

Squaring both sides, $P^2 = \frac{4\pi^2 a^3}{\mu}$. Since $\frac{4\pi^2}{\mu}$ is a constant, P^2 is proportional to a^3, and for earth satellites $P^2 = \left(2.805 \times 10^{-15} \frac{sec^2}{ft^3}\right)(a)^3$. Or, $P = \left(5.30 \times 10^{-8} \frac{sec}{ft^{3/2}}\right)(a)^{3/2}$

Problem: Initial data from Friendship 7 indicated that the booster burned out at a perigee altitude of 100 statute miles (SM), speed of 25,700 ft/sec, and flight path angle of 0°. Determine the speed and height at apogee, and the period (see fig. 2-17).

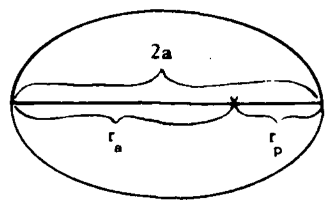


Figure 2-17. Orbit of friendship 7 (not to scale).

Given*: $h_p = 100 \text{ SM} = .5 \times 10^6$ ft $\qquad r_e = 20.9 \times 10^6$ ft

$v_{bo} = 25{,}700$ ft/sec

$\phi_{bo} = 0°$

Find: v_a, h_a, P

Solution:

$$E = \frac{v^2}{2} - \frac{\mu}{r} = \frac{(2.57 \times 10^4)^2}{2} - \frac{14.08 \times 10^{15}}{(20.9 + .5) \times 10^6}$$

$$= (3.31 \times 10^8) - (6.58 \times 10^8) = -3.27 \times 10^8$$

$$\text{but } E = -\frac{\mu}{2a};\ 2a = -\frac{\mu}{E} = \frac{-14.08 \times 10^{15}}{-3.27 \times 10^8} = 43.1 \times 10^6$$

From figure 2-17, $r_a + r_p = 2a$

$$\therefore r_a = 2a - r_p = (43.1 - 21.4) \times 10^6$$

$$= 21.7 \times 10^6$$

$$h_a = r_a - r_e = (21.7 - 20.9) + 10^6 = .8 \times 10^6 \text{ ft}$$

$$= 151 \text{ SM} \qquad \text{Answer}$$

*Even though $\phi_{bo} = 0$ if burnout altitude were not given as perigee altitude, you would have to determine if this were perigee or apogee. To do this, you would compute the circular speed for the given burnout altitude and compare this with the actual speed. If the circular speed is greater than the actual, burnout was at apogee; if the circular speed was less than the actual speed, burnout was at perigee.

$$H_p = H_a$$

$$v_p r_p = v_a r_a \qquad v_a = \frac{v_p r_p}{r_a}$$

$$v_a = \frac{(2.57 \times 10^4)(21.4 \times 10^6)}{21.7\ 10^6} = 25{,}400 \text{ ft/sec}$$

$$= 17{,}300 \text{ mph} \qquad \text{Answer}$$

$$P^2 = \frac{4\pi^2 a^3}{\mu} = (2.805 \times 10^{-15})(21.6 \times 10^6 \text{ ft})^3 \frac{\text{sec}^2}{\text{ft}^3} = 28.1 \times 10^6 \text{ sec}^2$$

$$P = 5.30 \times 10^3 \text{ sec} = 88.3 \text{ min} \qquad \text{Answer}$$

It is interesting to compare the computed apogee and period results with the actual orbit (later data gave a higher accuracy for burnout conditions):

Item	*Actual figures*	*Computed figures*
V_{bo}	25,728 ft/sec	25,700 ft/sec
h_{bo}	97.695 SM	100 SM
h_a	158.85 SM	151 SM
P	88.483 min	88.3 min

From the foregoing problem, it is evident that the principles and relatively simple algebraic expressions presented thus far are extremely important. They enable one to analyze a trajectory or orbit rather completely. We have confined the discussion to the two dimensional orbital plane. Before discussing some of the more interesting facets of orbital mechanics, it is necessary to properly locate a payload in three dimensions.

LOCATING BODIES IN SPACE

In one of the coordinate systems for space used by engineers and scientists, the origin is the center of the earth. This is a logical choice since the center of the earth is a focus for all earth orbits.

With the center of the coordinate system established, we must have a reference frame on which to make angular measurements with respect to the center. The reference frame should be regular in shape, and it should be fixed in space. A sphere satisfies the requirement of a regular shape. The sphere of the earth would be handy reference if it were fixed in space, but it rotates constantly. Therefore, we use the celestial sphere to satisfy the requirement for a reference frame. This is a nonrotating sphere of infinite radius whose center coincides with the center of the earth and whose surface contains the projection of the celestial bodies as they appear in the sky (fig. 2-18). The celestial equator is a projection of the earth's equator on the celestial sphere. We can project the track of a satellite on the celestial sphere by extending the plane of the orbit to its intersection with the celestial sphere.

After we have defined the center of the system and the celestial equator, we need a reference as a starting point for position measurements. We can find this point by passing a line from the center of the earth through the center of the sun to the celestial equator. This point, determined at the instant winter changes into spring, is the vernal equinox (fig. 2-19).

After we have established the references for the coordinates system, we must locate the orbit itself. The first item of importance is right ascension, (Ω), of the ascending node, which we define as the arc of the celestial equator measured eastward from the vernal equinox to the ascending node (fig. 2-18). The ascending node is the point where the projection of the satellite path crosses

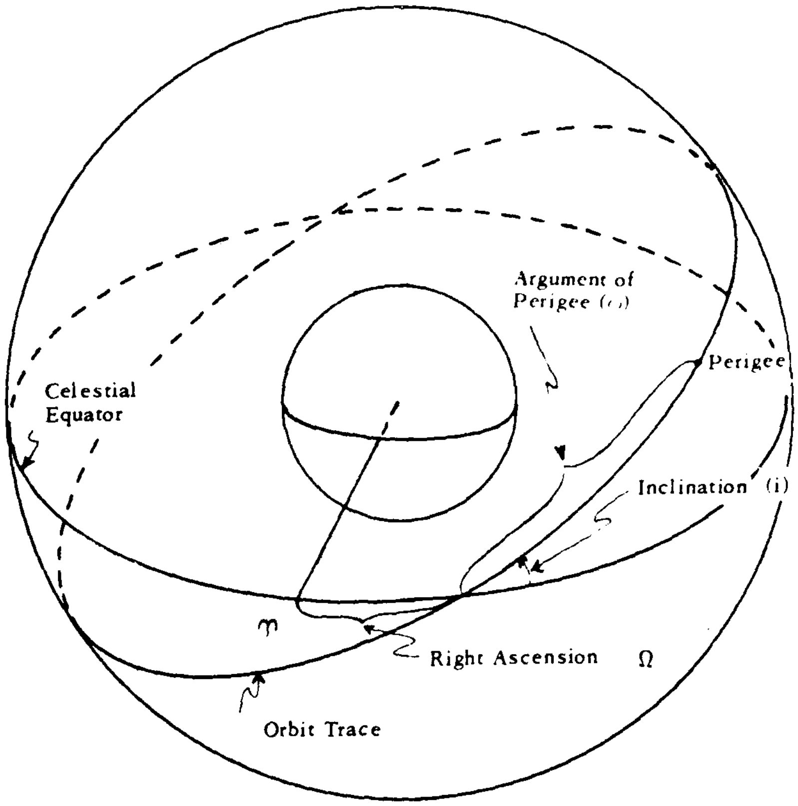


Figure 2-18. Celestial sphere.

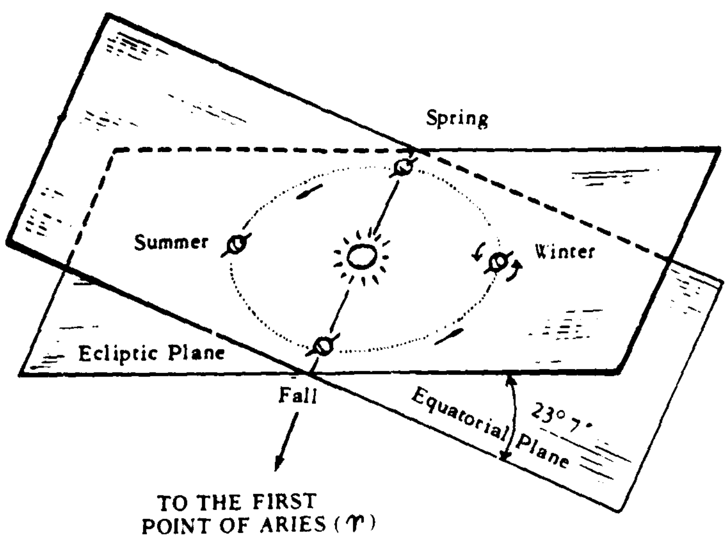


Figure 2-19. The vernal equinox.

the celestial equator from north to south. In other words, right ascension of the ascending node is the angle measured eastward from the first point of Aries to the point where the satellite crosses the equator from south to north.

The next item of importance is the angle the plane of the orbit makes with the equator. This is the angle of inclination, (i), which we define as the angle that the plane of the orbit makes with the plane of the equator, measured counterclockwise from the equator at the ascending node. *Equatorial* orbits have $i = 0°$; *posigrade* orbits have $i = 0°$ to $90°$; *polar* orbits have $i = 90°$; and *retrograde* orbits have $i = 90°$ to $180°$.

To describe the orbit further, we locate the perigee. The argument of perigee (ω) is the name for the angular measurement from the ascending node to the perigee, measured along the path of the orbit in the direction of motion.

If, in addition to the coordinates of orbit, we know a time of either perigee or right ascension of the node along with the eccentricity and the major axis of orbit, we can determine the exact position and velocity of the satellite at any time. Six quantities (right ascension, inclination, argument of perigee, eccentricity, major axis, and epoch time at either perigee or ascending node) form a convenient grouping of the minimum information necessary to describe the orbital path as well as the position of a satellite at any time. They constitute one set of orbital elements, known as the Set of Keplerian Elements.

Another interesting facet of earth satellites concerns the orbital plane. There is a relationship between the launch site and the possible orbital planes. This restriction arises from the fact that the center of the earth must be a focus of the orbit and must lie in the orbital plane.

Cos i (inclination) = cos (latitude) sin (azimuth), where the azimuth is the heading of the vehicle measured clockwise from true north, is the formula that determines the inclination of the orbital plane, i, to the equatorial plane.

As an example, a satellite launched from Cape Kennedy and injected at 30° N on a heading due east (azimuth 90°) will lie in an orbital plane that is inclined 30° to the equatorial plane.

$$
\begin{aligned}
\cos i &= (\cos \text{latitude})\,(\sin \text{azimuth}) \\
&= \cos 30° \sin 90° \\
\cos i &= \cos 30° \\
i &= 30°
\end{aligned}
$$

We can deduce from the preceding that the latitude of the launch site will define closely the minimum orbital plane inclination for a direct (no dogleg or maneuvering) injection. Thereby, all launch sites will permit direct injections at inclination angles from that minimum (the approximate latitude of the launch site) to polar orbits (plus retrograde supplements), provided there were no geographic restrictions on launch azimuth, such as range safety limitations. For example, direct injections from Vandenberg Air Force Base, California, (35° N) would permit inclination angles from about 35° to 145°. Once we define the inclination of the orbital plane, we can discuss the ground track.

SATELLITE GROUND TRACKS

The orbits of all satellites lie in planes that pass through the center of a theoretically spherical earth. Each plane intersects the surface of the earth in a great circle (fig. 2-20).

The intersection of the surface of the earth and a line between the center of the earth and a satellite forms a satellite's ground track. As the space vehicle moves in its orbit, this intersection traces out a path on the ground below. There are five primary factors that affect the ground track of a satellite moving along a free flight trajectory. These are:

1. Injection point
2. Inclination angle (i)
3. Period (P)
4. Eccentricity (e)
5. Argument of perigee (ω).

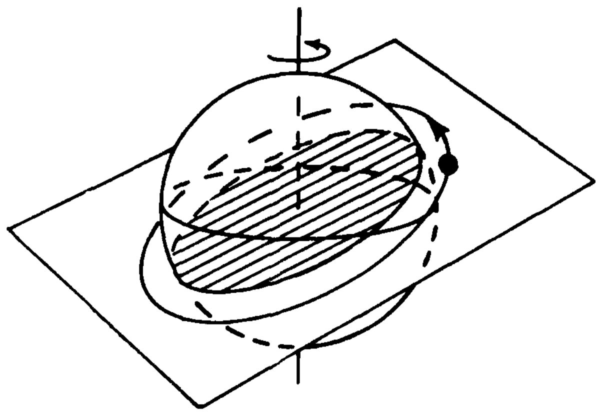


Figure 2-20. Satellite ground track geometry.

Of the preceding, the injection point simply determines the point on the surface from which the ground track begins, following orbital injection of the satellite. We have discussed the inclination angle in the previous section and will treat it below in further detail. Period, eccentricity, and argument of perigee each affect the ground track, but often difficulty results from efforts to isolate the effect of any one of the three. Therefore, we will make only general remarks regarding the three factors, rather than providing an intricate mathematical treatment.

If we predicate the study of satellite ground tracks upon a nonrotating earth, the track of a satellite in a circular orbit is easy to visualize. When the satellite's orbit is in the equatorial plane, the ground track coincides with the equator (fig. 2-21).

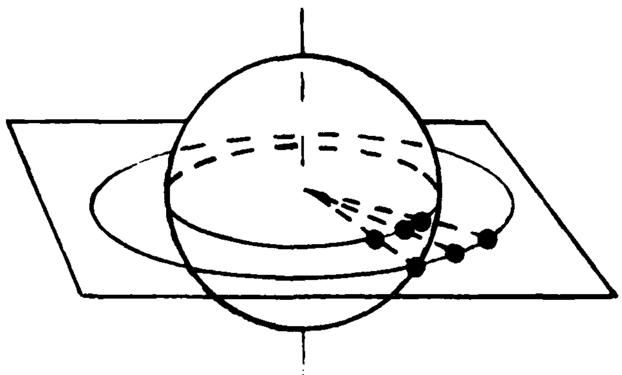


Figure 2-21. Equatorial track.

If the plane of the orbit is inclined to the equatorial plane, the ground track moves north and south of the equator. It moves between the limits of latitude equal to the inclination of its orbital plane (fig. 2-22). A satellite in either circular or elliptical orbit will trace out a path over the earth between these same limits of latitude, determined by the inclination angle. However, the satellite in elliptical orbit will, with one exception, remain north or south of the equator for unequal periods of time. This exception occurs when the major or long axis of the orbit lies in the equatorial plane.

Both the latitude of the vehicle and the direction of the vehicle's velocity at the time of injection or entry into orbit determine the inclination of an orbit. That is, the cosine of the inclination angle equals

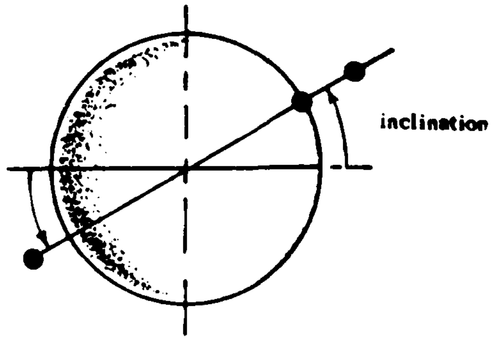


Figure 2-22. North-South travel limits.

the cosine of the latitude times the sine of the azimuth (when we measure it from the north). The minimum inclination that an orbital plane can assume is the number of degrees of latitude at which injection occurs. This minimum inclination occurs when the direction of the vehicle's velocity is due east or west at the time of injection. If the vehicle's direction at injection into orbit is any direction other than east or west, the inclination of the orbital plane will increase (fig. 2-23).

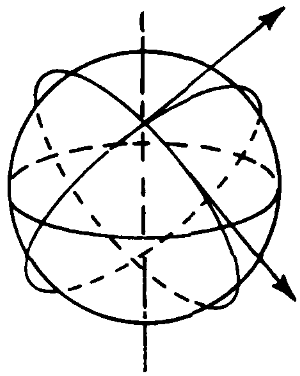


Figure 2-23. Injection-inclination geometry.

On a flat map of the earth, satellite ground tracks appear to have different shapes than on a sphere. The ground track for a vehicle in an inclined circular or elliptical orbit appears as a sinusoidal trace with north-south limits equal to the inclination of the orbital plane (fig. 2-24).

When we consider the earth's rotation, visualizing a satellite's ground track becomes more complex. A point on the equator moves from west to east more rapidly than do points north and south of the equator. Their speeds are the speed of a point on the equator times the cosine of their latitude. Satellites in circular orbit travel at a constant speed. However, when the orbit is inclined to the equator, the component of satellite velocity that is effective in an easterly or westerly direction varies continuously throughout the orbital trace (fig. 2-25). As the satellite crosses the equator, its easterly or westerly component of velocity is its instantaneous total velocity times the cosine of its angle of inclination. When it is at the most northerly or southerly portion of its orbit, its easterly or westerly component equals its total instantaneous velocity.

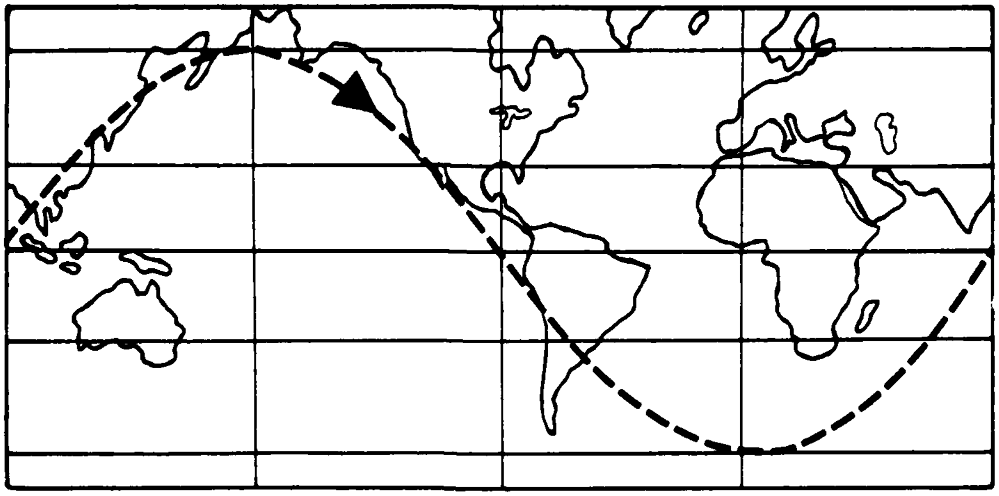


Figure 2-24. Ground track on flat, nonrotating earth map.

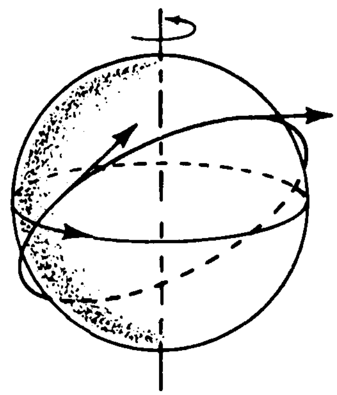


Figure 2-25. Effective east/west component of satellite velocity.

In elliptical orbits only the horizontal velocity component contributes to the satellite's ground track. Further complication results because the inertial or absolute speed of the satellite varies throughout the elliptical path (fig. 2-26).

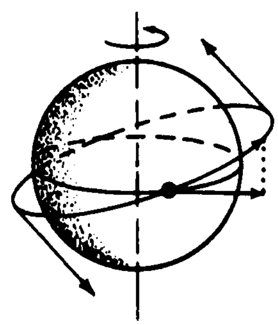


Figure 2-26. Variation of velocity magnitude in an elliptical orbit.

Because the ground track is dependent upon the relative motion between the satellite and the earth, the visualization of ground tracks becomes quite complicated. Earth rotation causes each successive track of a satellite in a near-earth orbit to cross the equator west of the preceding track (fig. 2-27). This westerly regression equals the period of the satellite times the rotational speed of the earth. We can see the regression more clearly if we consider the angular speed. The earth's angular speed of rotation is 15° per hour. We can determine the number of degrees of regression (in terms of a shift in longitude) by multiplying the period of the satellite by 15° per hour, the angular speed of the earth. If we increase the altitude of a satellite, thereby increasing the time required to complete one revolution in the orbit, the distance between successive crossings of the equator increases. Again, using a flat map of the earth, the track of a satellite in circular orbit (with a period of less than 20 hrs) appears as a series of sinusoidal traces, each successively displaced to the west (fig. 2-28).

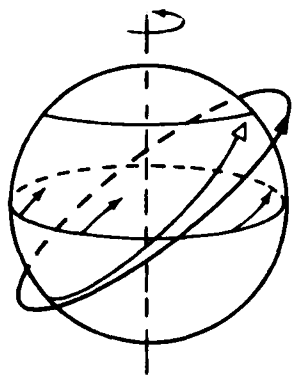


Figure 2-27. Ground track regression due to earth rotation.

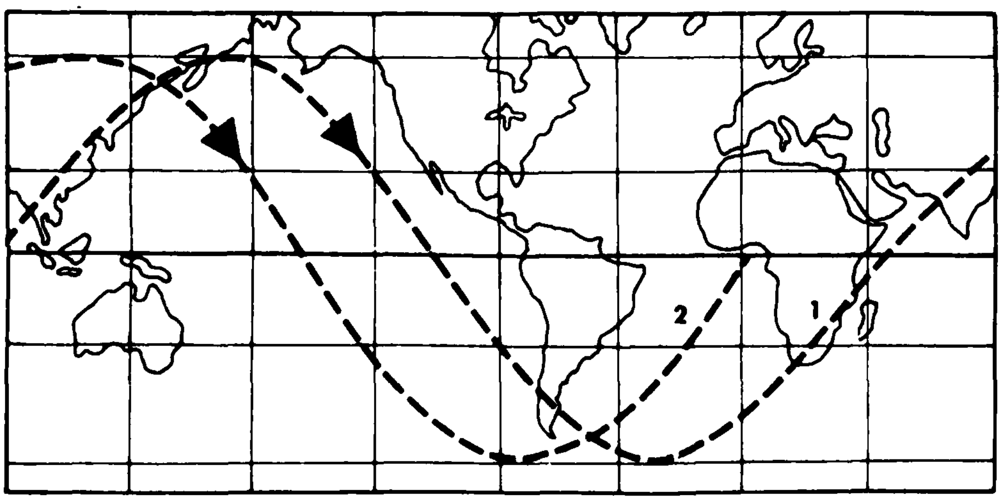


Figure 2-28. Westward regression of sinusoidal tracks.

The ground track of a satellite in elliptical orbit results in a series of irregular traces on a flat map which have one lobe larger than the other. Orbital time determines the amount of compression that occurs to these lobes, a combination of factors (inclination, eccentricity, period, and location of

perigee) alter the shape of the lobes, and successive traces are displaced to the west (fig. 2-29).

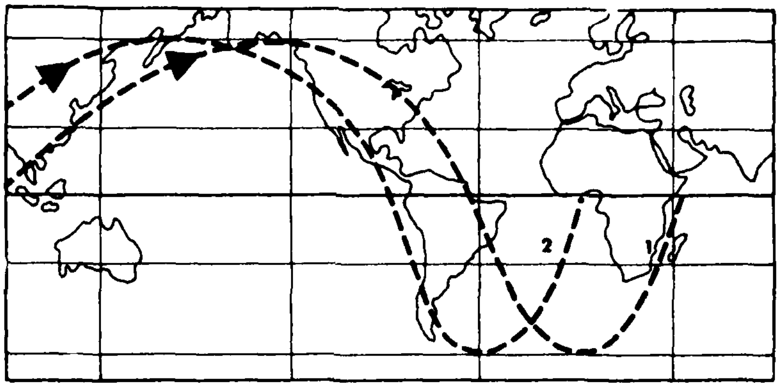


Figure 2-29. Westward regression of irregular tracks.

Some satellites follow orbits that have particularly interesting ground tracks. A satellite with a 24-hour period of revolution is one such case. If this satellite is in a circular orbit in the equatorial plane, we refer to it as a synchronous satellite; its trace is a single point. If it orbits in the polar plane, it will complete half of its orbit while the earth is rotating halfway about its axis. The result is a trace that crosses a single point on the equator as the satellite crosses the equator heading north and south. The complete ground track forms a figure eight. If the plane of the circular orbit is inclined at smaller angles to the equator, the figure eights are correspondingly smaller (fig. 2-30).

We can alter the shape of the figure eight by placing a satellite in an elliptical flight path. The eccentricity of the ellipse changes the relative size of the loops of the figure eight. If we fix eccentricity and inclination, then changing the location of perigee will vary the shape and orientation of the figure eight as shown in figure 2-31. The conditions and geographical location of injection into the 24-hour orbit determine the longitude of perigee. However, the inclination of the orbital plane, (i), and the argument of perigee, (ω), fix the latitude of perigee:

$$\text{Sin (latitude of perigee)} = \sin i \sin \omega$$

Apogee is on the same meridian as perigee and at the same degree of latitude but in the opposite hemisphere.

Certain navigation satellite concepts have proposed circular ground tracks similar to the one shown in figure 2-31a(g). In this case, with apogee in the northern hemisphere, we might use a multiple satellite system using this type of ground track for supersonic aircraft navigation in the North Atlantic.

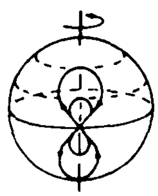
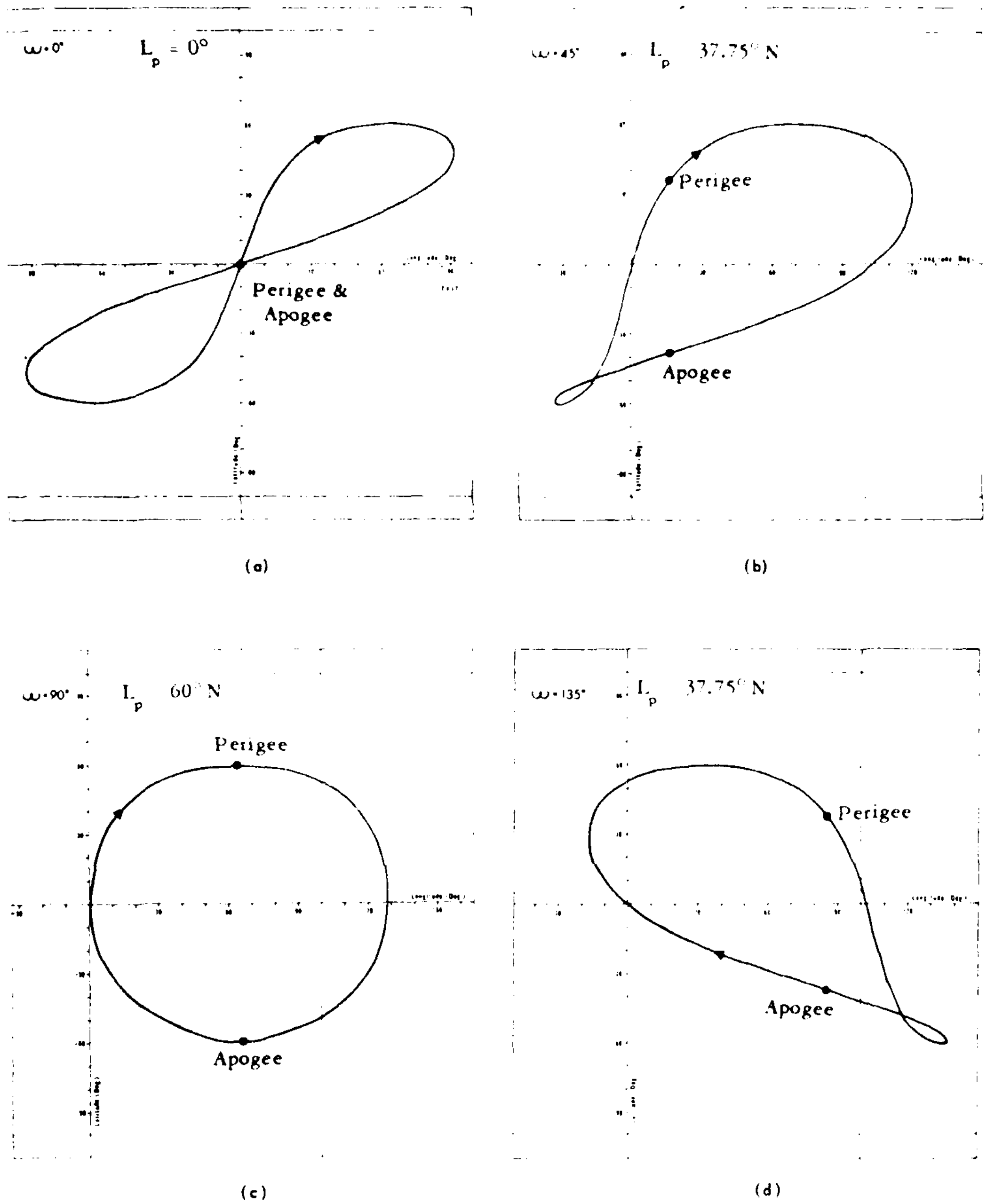
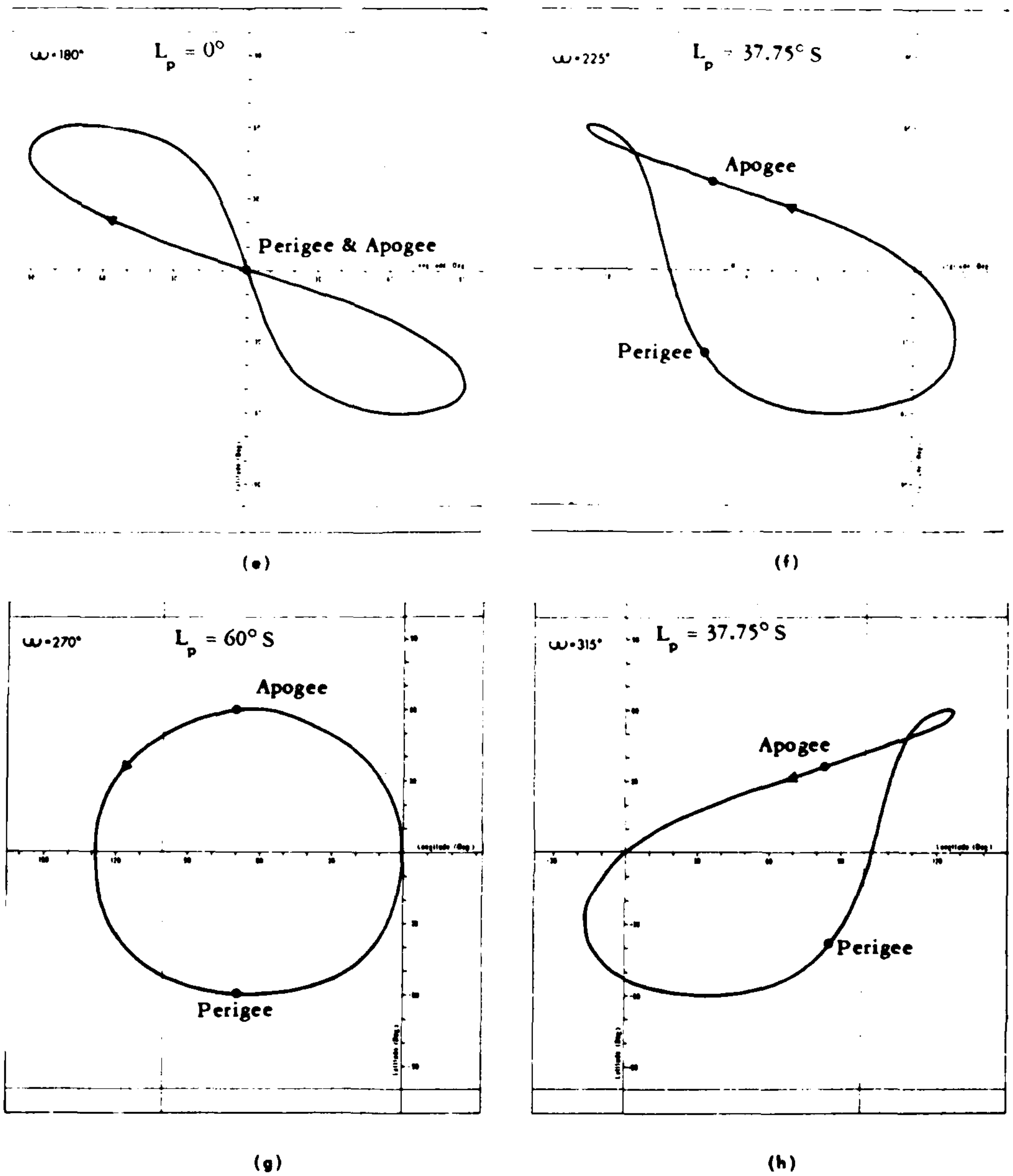


Figure 2-30. Figure eights for inclinations less than polar.



Eccentricity (ε) = .6
Inclination (i) = 60°
Argument of Perigee (ω) = As Stated

Figure 2-31. Variation of elliptical, 24-hour track with movement of perigee



Eccentricity (ϵ) = .6
Inclination (i) = 60°
Argument of Perigee (ω) = As Stated

Figure 2-31a. Variation of elliptical, 24 hour track with movement of perigee, continued.

If a satellite is in an orbit with a period much greater than the earth's 24-hour period of rotation, the satellite appears as a point in space under which the earth rotates. If the orbit is in the equatorial plane, the trace is a line on the equator moving to the west. If the spacecraft orbit is inclined to the equator, its earth track will appear to be a continuous trace wound around the earth like a spiral between latitude limits equal to its angle of inclination (fig. 2-32). An example of this phenomenon is the ground track of the moon.

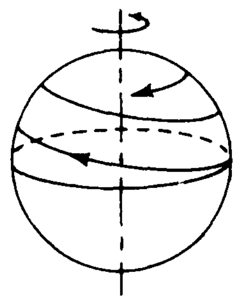


Figure 2-32. Track for inclined orbit with period greater than one day.

It should be clear that there is an almost limitless variety of satellite ground tracks. To obtain a particular track, it is only necessary to select the proper orbit. If changes are made in the inclination of an orbital plane to the equator, if its period is varied, if the eccentricity is controlled, or, if the location or perigee is specified, many different ground tracks can be achieved.

SPACE MANEUVERS

One characteristic of satellites is that their orbits are basically stable in inertial space. This stability is often an advantage, but it can pose problems. Space operations such as resupply, rendezvous, and interception may require that we change the orbits of space vehicles. Such changes are usually changes in orbital altitude, orbital plane (inclination), or both. In this section we will review some methods of maneuvering in space.

Altitude Change

When a satellite or space vehicle changes its orbit in altitude, it requires additional energy. This is true whether it increases or decreases the altitude. The classic example of changing the orbital altitude of a satellite is the Hohmann transfer.

The Hohmann transfer is a two-impulse maneuver between two circular, coplanar orbits. For most practical problems, this method uses the least amount of fuel and is known as a minimum energy transfer. The path of the transfer follows an ellipse that is cotangential to the two circular orbits (fig. 2-33).

A Hohmann transfer requires two applications of thrust. Each application of thrust changes the speed of the vehicle and places it into a new orbit. Obviously, we must control accurately both direction and magnitude of the velocity change, Δv, for a precise maneuver. If we desire an increase in altitude, the point of departure becomes the perigee of the transfer ellipse; the point of interjection to the higher circular orbit becomes the apogee of the transfer ellipse. (For the transfer ellipse, $2a = r_1 + r_2$.) To lower altitude, the reverse is true. The point of departure will be the apogee of the transfer ellipse.

In general, we can divide the process for determining the total increment of velocity, Δv, required to complete a Hohmann transfer into seven steps.

1. Determine the velocity the vehicle has in the initial orbit.

$$V_{c1} = \sqrt{\frac{\mu}{r_1}}$$

2. Determine the velocity required at the initial point in the transfer orbit.

$$v_1 = \sqrt{\frac{2\mu}{r_1} - \frac{\mu}{a}}$$

3. Solve for the vector difference between the velocities found in steps 1 and 2.

4. Find the velocity the vehicle has at the final point in the transfer ellipse.

$$v_2 = \sqrt{\frac{2\mu}{r_2} - \frac{\mu}{a}} \text{ or } v_2 = \frac{v_1 r_1}{r_2}$$

5. Compute the velocity required to keep the vehicle in the final orbit.

$$V_{c2} = \sqrt{\frac{\mu}{r_2}}$$

6. Find the vector difference between the velocities found in steps 4 and 5.

7. Find the total Δv for the maneuver by adding the Δv from step 3 to the Δv from step 6.

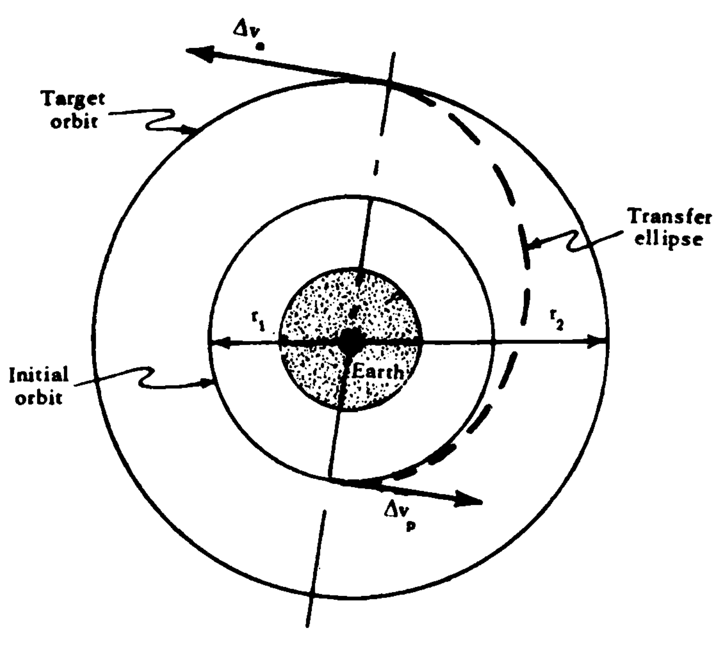


Figure 2-33.

From a practical point, once we know the required Δv, we can compute the amount of propellant required for the maneuver from $\frac{\Delta v}{I_{sp} g} = \ln\left(\frac{W_o}{W}\right) = \ln$ mass ratio, as discussed in chapter 3.

There are other ways to accomplish an altitude change. One such method, the fast transfer, is useful when time is a factor. In this method, the transfer ellipse is not cotangential to the final orbit but crosses it at an angle (fig. 2-34). Again, we apply the Δv in two increments, but Δv_2, applied at the intersection of the transfer ellipse and the target orbit, must achieve the desired final velocity in the proper direction. The steps for calculating the required Δv are similar to those for the Hohmann transfer, noting that we must treat velocity differences as vectorial quantities. For the same altitude change the fast transfer requires more Δv.

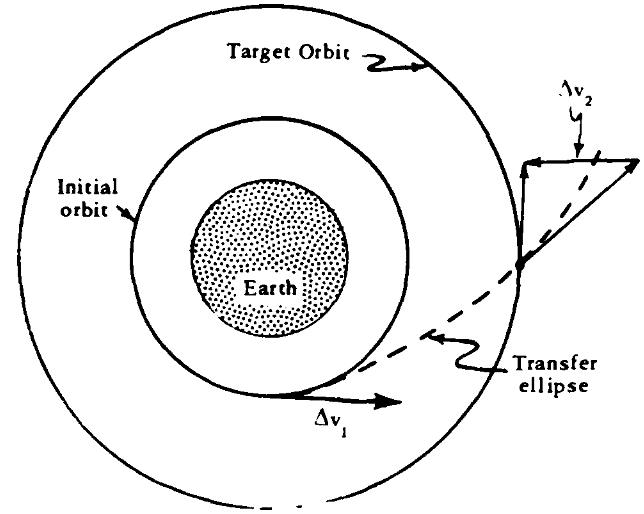


Figure 2-34.

Other texts on astronautics discuss other methods of changing altitudes. The procedures are similar to those discussed here. The magnitude and direction of the velocity vector remain the critical factors.

Plane Change

Changing the orbital plane of a satellite requires the expenditure of energy. This is apparent if we examine the vector diagram representing two circular orbits of the same altitude, the only difference being in their inclinations (fig. 2-35).

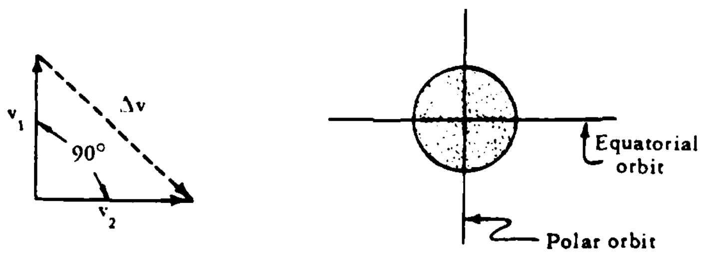
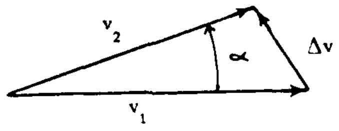


Figure 2-35.

In this case, the orbital speeds, v_1 and v_2, are equal except that they are 90° to each other. Transferring from the polar orbit to the equatorial orbit would require the Δv represented by the dashed arrow. For a 90° plane change (an extreme case) the Δv exceeds the existing orbital speed. We should note that we can accomplish a change from one plane to another only at the intersection of the two planes.

We can determine the Δv required to change the orbital plane any specified amount by examining the vectors involved. The problem is solvable by use of the Law of Cosines. For example, if v_1 represents the existing orbital velocity, v_2 the final orbital velocity, and α the desired plane change angle,* then the formula in figure 2-36 results.



$$\Delta v^2 = v_1^2 + v_2^2 - 2v_1 v_2 \cos \alpha$$

Figure 2-36.

If we are to change only the plane, then $v_1 = v_2$ and the problem is simplified. But, if we change the altitude (or eccentricity) as well, v_1 and v_2 will not be equal.

The amount of Δv required to accomplish a plane change depends on the amount of change desired. It is a function of the altitude at which we make the change.

It requires less Δv to make a plane change at high altitudes than at low altitudes because the orbital speed of the vehicle is less at higher altitudes. In other words, it is more economical, in terms of propellant required, to make plane changes where the speed of the satellite is low—at apogee, or at high altitudes.

Combined Maneuvers

If a requirement exists to perform both a plane change and an altitude change, some economy will result if we combine the operations. We can solve the problem of combining a plane and altitude change quite simply by considering the vector diagram. For example, we want to change the altitude of a vehicle from 100 nautical mile circular orbit to 1,500 nautical mile circular orbit with a plane change of 10°. Recognizing that we can make the plane change more economically at altitude, we plan to combine the plane change with injection from the Hohmann transfer ellipse into the 1,500 nautical mile circular orbit. We initiate a typical Hohmann alaitude change at the point of intersection of the two planes by increasing the vehicle's speed.

At the apogee of the transfer ellipse (1,500 nautical mile altitude) the vehicle's speed is 19,800 ft/sec. The required circular speed is 21,650 ft/sec. The complete problem, at apogee, is seen in figure 2-37.

We calculate the Δv required for the combined maneuver by use of the Law of Cosines. This is: $\Delta v =$ 4,055 ft/sec. It is necessary to compute the angle at which we apply the Δv. We can accomplish this calculation with the Law of Sines.

* See appendix B.

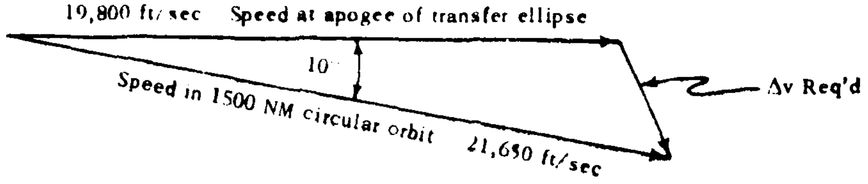


Figure 2-37.

PERTURBATIONS

When we use the assumptions given thus far to calculate the orbit of an artificial satellite, it will vary slightly from the actual orbit unless we make corrections to take care of outside forces. These outside forces, known as perturbations, cause deviations in the orbit from those predicted by two-body orbital mechanics. To get a better understanding of how the satellite's actual orbit is going to behave, we must consider the following additional factors:

1. The earth is not the only source of gravitational attraction on the satellite since there are other gravitational fields (principally of the sun and moon) in space. This effect is greatest at high altitudes (above 20,000 nautical miles).
2. The earth is not a spherically homogeneous mass but has a bulge around the equatorial region. This additional mass causes the gravitational pull on the satellite *not* to be directed toward the center of the earth. This is the major perturbation effect at medium altitudes (between 300 nautical miles and 20,000 nautical miles).
3. The earth has an atmosphere that causes drag. This effect is most significant at low altitudes (below 300 nautical miles).

Third Body Effects

One cause of perturbations is the introduction of one or more additional bodies to the system creating a problem involving three or more bodies. Figure 2-38 shows an exaggerated perturbation, referred to as a hyperbolic encounter. Initially, the satellite is in orbit 1 about an attracting body such as the earth. As the satellite approaches the moon, the gravitational influence of the moon dominates, and the center of the moon becomes the focus instead of the center of the earth. Since the vehicle approaches the moon with more than escape velocity, it must leave the moon's sphere of gravitational influence with greater than escape velocity. This means that the vehicle's velocity with respect to the moon is greater than that required to escape. Therefore, for the short time that the moon is the attracting body, the vehicle is on a hyperbolic path with respect to the moon. When the satellite leaves the sphere of influence of the moon, it switches back to the earth's sphere of influence and goes into a new elliptical orbit about the earth. The next time the satellite returns to this region, the moon will have moved in its orbit, and the satellite will maintain orbit 2. A hyperbolic encounter is a method of changing the energy level of a satellite. By proper positioning, we could use a hyperbolic encounter to increase or decrease the energy of a space vehicle. The change in energy of the second body offsets the change in energy of the vehicle (the moon in the case illustrated).

Effects of Oblate Earth

Another cause of perturbations is the bulge of the earth at the equator sometimes called the earth's oblateness. We can see the effect of this oblateness on the satellite if we imagine that a sphere which has an added belt of mass wrapped around the equatorial region makes up the earth. As shown in figure 2-39 the much smaller but significant attractions F_1 and F_2 directed toward the near and far sides of the equatorial bulge "disturb" the primary gravitational attraction F, directed to the center of the earth. With r_1 smaller than r_2, F_1 will be larger than F_2, and the resultant force obtained by

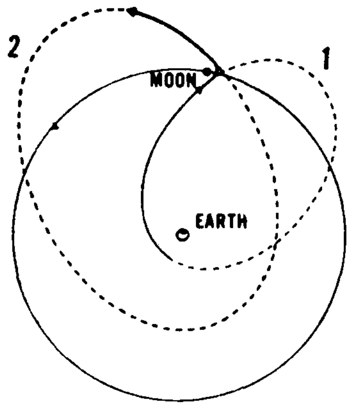


Figure 2-38. Hyperbolic encounter.

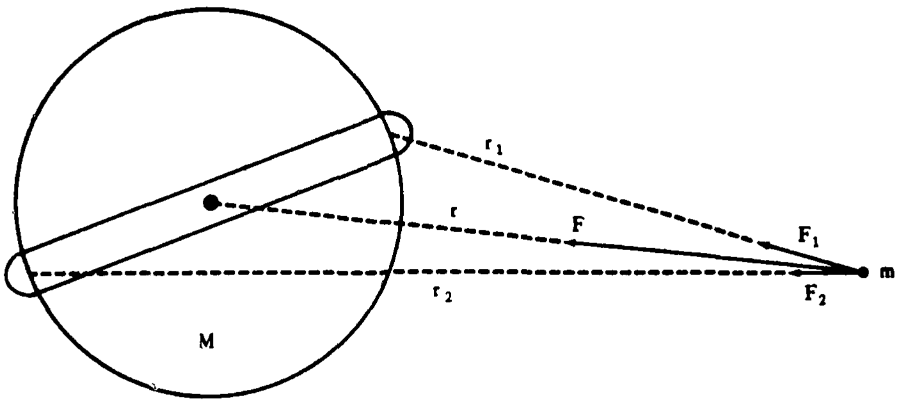


Figure 2-39. Deviation in force vector caused by oblateness of the earth.

combining F, F_1, and F_2 will point no longer to the center of the earth but will be deflected slightly toward the equator on the near side. As the satellite moves in its orbit the amount of this deflection will change depending on the vehicle's relative position and proximity to the equatorial region. Two perturbations that result from this shift in the gravitational forces are regression of the nodes and rotation of the line of apsides (major axis) or rotation of perigee.

Figure 2-40 illustrates regression of the nodes as a rotation of the plane of the orbit in space. The resulting effect is that the nodes, both ascending and descending, move west or east along the equator with each succeeding pass. The direction of this movement will be opposite to the east or west component of the satellite's motion. Satellites in the posigrade orbit (inclinations less than 90°), illustrated in figure 2-40, have easterly components of velocity so that the nodal regression in this case is to the west. Retrograde orbits with their westerly component of velocity will reverse the movement of the nodes. Figure 2-41 shows why, for a vehicle traveling west to east, the regression of the nodes is toward the west. The original track from A to B would cross the equator at Ω_1. Simplifying the effect of the equatorial bulge to a single impulse at point E, the spacecraft moves the track so that it crosses the equator at Ω_2. At point F the simplified effect of the bulge is a single impulse down, changing the orbital path line to the line FD, which would have crossed the equator at Ω_3. This effect, regression of the nodes, is more pronounced on low-altitude satellites than high-altitude satellites. In low-altitude,

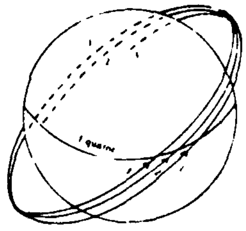


Figure 2-40. Regression of the nodes.

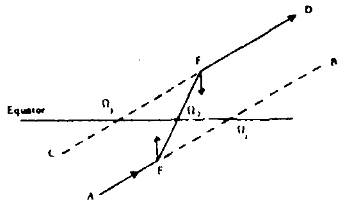


Figure 2-41. Regression of nodes toward the west when vehicle is traveling west to east.

low-inclination orbits the regression rate may be as high as 9° per day. Figure 2-42 shows how the regression rate changes for circular orbits at various altitudes and inclination angles. Note that nodal regression is zero in the polar orbit case. It has no meaning in equatorial orbits.

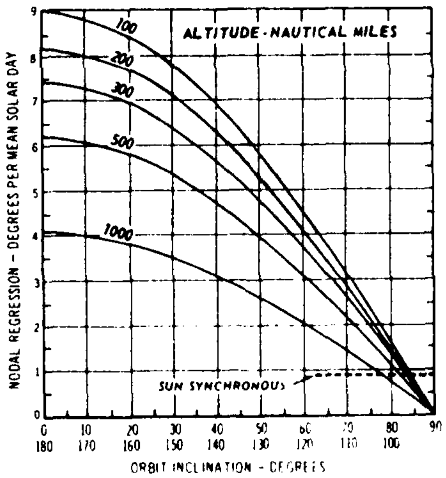


Figure 2-42. Nodal regression rate per day for circular orbits.

Satellites requiring sun synchronous orbits (for photography or other reasons) are an example of how we can use regression of the nodes to practical advantage in certain situations. Figure 2-43 shows the injection of the satellite into an orbit passing over the equator on the sunlit side of the earth at local noon (the sun overhead). This condition initially aligns the orbital plane so that it

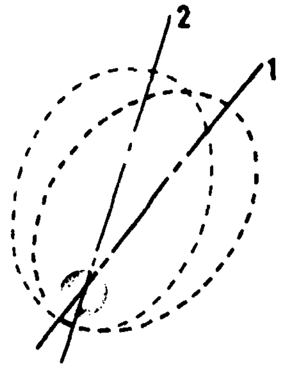


Figure 2-44. Earth's equatorial bulge changes the argument of perigee.

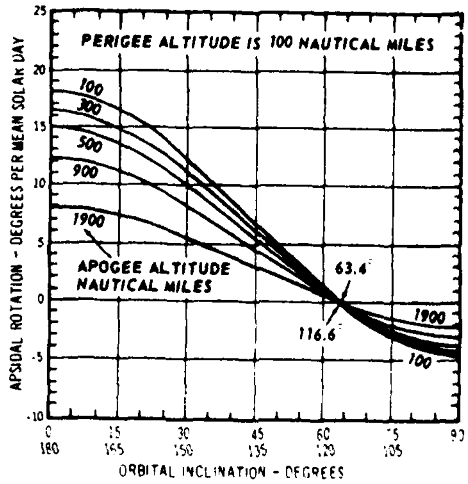


Figure 2-45. Apsidal rotation rate per day for orbits with 100 NM perigee altitude.

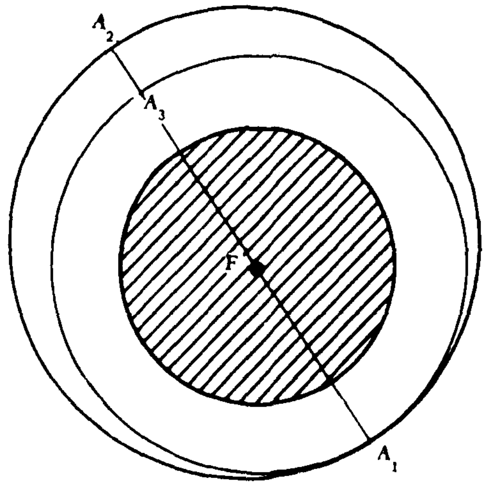


Figure 2-46. Decrease in the eccentricity of a satellite orbit cased by drag.

3. **Specific Mechanical Energy:**

$$E = \frac{v^2}{2} - \frac{\mu}{r}$$

$$E = -\frac{\mu}{2a}$$

4. **Specific Angular Momentum:**

$$H = v\,r\cos\phi$$

5. **Two Body Relationships:**

$$V_{circle} = \sqrt{\frac{\mu}{r}}$$

$$V_{ellipse} = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}$$

$$V_{escape} = \sqrt{\frac{2\mu}{r}} = \sqrt{2gr}$$

$$P = \frac{2\pi a^{\frac{3}{2}}}{\mu^{\frac{1}{2}}} = \left(5.30 \times 10^{-8} \frac{\text{sec}}{\text{ft}^{\frac{3}{2}}}\right)(a)^{\frac{3}{2}}$$

$$P^2 = \frac{4\pi^2 a^3}{\mu} = \left(2.805 \times 10^{-15} \frac{\text{sec}^2}{\text{ft}^3}\right)(a)^3$$

6. **Variation of g:**

$$g = \frac{\mu}{r^2}$$

7. **Law of Cosines:**

$$a^2 = b^2 + c^2 - 2bc\cos A$$

Law of Sines:

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

8. **Constants:**

$$r_e = 20.9 \times 10^6 \text{ ft}$$

$$r_e = 3{,}440 \text{ NM}$$

$$\mu = 14.08 \times 10^{15} \frac{\text{ft}^3}{\text{sec}^2} \text{ for earth.}$$

$$1 \text{ NM} = 6{,}080 \text{ ft}$$

$$\pi \text{ radians} = 180^\circ$$

$$1 \text{ radian} = 57.3^\circ$$

9. **Inclination of Orbital Plane**

$$\cos i = \cos \text{Lat} \sin \text{Azimuth}$$

BIBLIOGRAPHY

Berman, Arthur I. *The Physical Principles of Astronautics.* New York: John Wiley and Sons, Inc., 1961.

Department of Astronautics, US Air Force Academy. *Astronautics, Elements of Free Flight Trajectories.* US Air Force Academy, 1961.

Ehricke, Kraft. *Environment and Space Mechanics.* Vol. 1 of *Space Flight.* Princeton, N.J.: Van Nostrand and Co., 1960.

________________. *Dynamics.* Vol. 2 of *Space Flight.* Princeton, N.J.: Van Nostrand and Co., 1960.

Ely, Lawrence D. *Space Science for .he Layman.* Springfield, Ill.: Charles C. Thomas Co., 1967.

El'Yasberg, P.E. *Introduction to the Theory of Flight of Artificial Earth Satellites.* (Translated from Russian.) Jerusalem: Israel Program for Scientific Translations, 1967.

Fowles, Grant R. *Analytical Mechanics.* Chicago: Holt, Rinehart, and Winston, 1963.

Glasstone, Samuel. *Sourcebook on the Space Sciences.* Princeton, N.J.: Van Nostrand Co., 1965.

Koestler, Arthur. *The Sleepwalkers.* New York: The MacMillan Co., 1959.

LeGalley, Donald P. *Guidance, Navigation, Tracking, and Space Physics.* Vol. 3 of *Ballistic Missile and Space Technology.* New York: Academic Press, 1960.

Ruppe, Harry O. *Introduction to Astronautics.* New York: Academic Press, 1966.

Szebehely, Victor G. *Celestial Mechanics and Astrodynamics.* Vol. 14 of the *Progress in A· ·nautics and Aeronautics.* New York: Academic Press, 1964.

________________. *Theory of Orbits.* New York: Academic Press, 1967.

US Naval Academy. *Space Flight Handbook.* Washington D.C.: US Government Printing Office, 1962. (NAVPERS 92988).

Chapter 3

PROPULSION SYSTEMS

Throughout history, the methods of propulsion that people have been able to control have allowed them to go faster, farther, and higher. Today they use rockets to send payloads into space. A space mission needs a launch vehicle or booster that will withstand high acceleration and aerodynamic forces as it lifts the payload from the surface of the earth. To be successful, the mission requires adequate ground equipment to launch and track the flight vehicle, propulsion and guidance to place it on the desired trajectory, reliable electric power sources, and communication equipment to send data back to earth. Finally, mission success requires trained personnel who follow correct procedures to ensure that all subsystems operate properly.

Rocket propulsion is vital in any successful space program. WIthout it, there would be no payloads in space. Because the scope of the rocket propulsion field is far too extensive, we cannot treat it comprehensively in a single chapter. Therefore, we cover only a limited number of topics in this chapter. These topics include the theory of rocket propulsion, rocket propellants, types of chemical rocket engines, and advanced propulsion techniques.

THEORY OF ROCKET PROPULSION

Newton's three laws of motion apply to all rocket-propelled vehicles. They apply to gas jets used for attitude control, small rockets used for stage separation or for trajectory correction, and large rockets used to launch a vehicle from the surface of the earth. They apply to nuclear, electric, and other advanced types of rockets, as well as to chemical rockets. Newton's laws of motion are stated briefly as follows:

1. Bodies in uniform motion, or at rest, remain so unless acted upon by an external unbalanced force.
2. The force required to accelerate a body is proportional to the product of the mass of the body and the acceleration desired.
3. To every action there is an equal and opposite reaction.

In relating these laws to propulsion, we may paraphase and simplify them. For example, the first law says, in effect, that the engines must be adequate to overcome the inertia of the launch vehicle. The engines must be able to start the vehicle moving and accelerate it to the desired velocity. Another way of expressing this for a vertical launch is to say that the engines must develop more pounds of thrust than the vehicle weighs. Some space missions require slowing the vehicle or changing its direction. We must apply an unbalanced force to accomplish these tasks. Newton's first law is universal and therefore applies not only to a vehicle at rest on a planet, but also to a vehicle in the so-called "weightless" condition of free flight.

We must consider several forces when applying the second law. For example, the accelerating force is the net force acting on the vehicle. This means if we launch a 100,000-pound vehicle vertically from the earth with a 150,000-pound thrust engine, there is a net force at launch of 50,000 pounds—the

difference between engine thrust and vehicle weight. Here the force of gravity is acting opposite to the direction of the thrust of the engine.

Propellants comprise approximately 90 percent of the vehicle's weight at launch. The engines expend propellants as they run, decreasing the vehicle weight. Therefore, the net force acting on the vehicle increases, and the vehicle accelerates rapidly.

Figure 3-1 shows the acceleration and the resulting velocities attained during powered flight. The acceleration and velocity are low at launch and just after launch due to the small net force acting at that time. Both acceleration and velocity increase rapidly as the engine burns the propellants. When the vehicle shuts off the first stage engine and staging occurs, acceleration drops sharply. When the second stage engine ignites, acceleration and velocity will increase again. As the upper stage rocket engines burn more propellants, rapid increases in acceleration and velocity occur. When the vehicle reaches the correct velocity (speed and direction) and altitude for the mission, it terminates the thrust. Acceleration drops to zero after thrust termination, or burnout, and the vehicle begins free flight. For vehicles with three, four, or more stages, similar changes appear in both the acceleration and velocity each time staging occurs. Staging a vehicle increases the velocity in steps to the high values required for space missions.

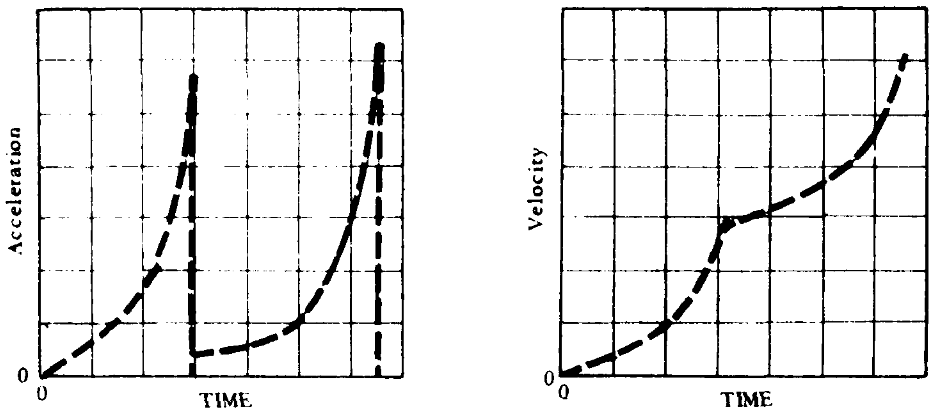


Figure 3-1. Powered flight of a typical two-stage rocket.

Newton's second law also applies to a vehicle in orbit. Since the vehicle is in a "weightless" condition, the acceleration for a given thrust depends on the vehicle mass, because the vehicle still has mass even though it is "weightless." Remember that mass is the quantity of matter in a body, whereas weight is the force exerted on a given mass by a gravitational field. The mass of a body is the same everywhere, but its weight depends on its mass, the gravitational field, and its position in the gravitational field.

A given thrust will give a large acceleration to a vehicle with a small mass or a small acceleration to a vehicle with a large mass. Even a very small thrust (0.1 pound or less) operating for a long time can accelerate an earth-orbiting vehicle to the velocities needed to go to the other planets of our Solar System. The section of this chapter on low-thrust engines discusses such propulsion systems.

To relate Newton's *third*, or "action-reaction," *law* to rocket propulsion, consider what happens in the rocket motor (fig. 3-2). All rockets develop thrust by expelling particles (mass) at high velocity from their exhaust nozzles. The effect of the ejected exhaust appears as a reaction force, called thrust, acting in a direction opposite to the direction of the exhaust.

Today, practically all exhaust nozzles, whether for a liquid- or a solid-propellant rocket, use some form of the de Laval (converging-diverging) nozzle. It accelerates the exhaust products to supersonic velocities by converting some of the thermal energy of the hot gases into kinetic energy.

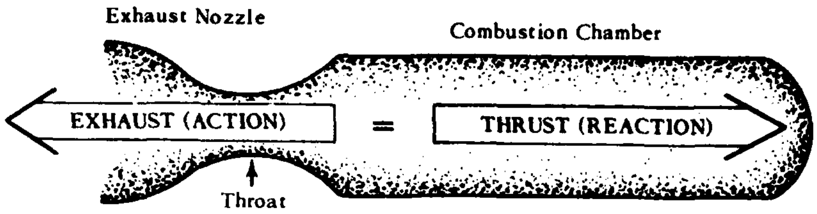


Figure 3-2. Action-reaction in a rocket motor.

In the combustion chamber, the burning propellants have negligible unidirectional velocity but have high temperature and pressure. The high pressure forces the gases through the nozzle to the lower pressure outside the rocket. As the gases move through the *converging* section of the nozzle, their temperature and pressure decrease, and their velocity increases to Mach 1 (the speed of sound) at the smallest cross-sectional area of the nozzle, which we call the throat. The gases will attain sonic velocity at the nozzle throat if the combustion chamber pressure is approximately twice the throat pressure.

Since the speed of sound increases with an increase in the temperature of the propagating gas, both sonic velocity and actual linear gas velocity in the throat increase with an increase in gas temperature. Therefore, the higher the gas temperature at the throat, the higher the exhaust velocities that are generated.

Thrust

Thrust (F) of a rocket is a sum of two terms, "momentum thrust" and "pressure thrust." Momentum thrust is the product of the *propellant rate of flow* and the *velocity of the exhaust* relative to the rocket. Pressure thrust is the product of the maximum cross-sectional area of the divergent nozzle section and difference between *exit pressure* of the exhaust and the *ambient pressure* surrounding the rocket. The functional thrust equation below shows these relationships:

$$F = \underbrace{\frac{\dot{W}}{g} v_e}_{\text{Momentum Thrust}} + \underbrace{A_e (P_e - P_o)}_{\text{Pressure Thrust}} \qquad (1)$$

In the above equation:

F = Thrust developed (lb_f)
$\dot{W}$ = Mass rate flor of propellants (lb_m/sec)
V_e = Velocity of gases at nozzle exit (ft/sec)
A_e = Cross-sectional area of nozzle exit (in^2)
P_e = Pressure of gases at nozzle exit (lb/in^2)
P_o = Ambient pressure (lb_f/in^2)

The term "g_c" is a constant to convert the units used in the archaic British engineering system. By definition, one pound of *force* will accelerate one pound of *mass* at 32.174 feet per second, or:

$$1\ lb_f = 32.174\ lb_m - ft/sec^2,$$

$$g_c = 32.174 \frac{lb_m - ft}{lb_f - sec^2} = \text{unity}$$

This conversion constant has nothing to do with the local force of gravity. Note the distinction between the pound-mass unit and the pound-force unit.

In high-thrust rockets, which eject many pounds of propellant (or exhaust products) per second at velocities of several thousands of feet per second, the *momentum thrust is by far the dominant part* of the total thrust. It usually comprises more than 80 percent of the thrust being developed.

Nozzles and Expansion Ratio

The condition of optimum expansion occurs in a rocket nozzle when the pressure of the exhaust gases at the nozzle exit (P_e) equals the ambient (atmospheric) pressure (P_o). When $P_e = P_o$, the thrust output of the engine is the maximum that particular engine can obtain at that altitude. For any given nozzle this condition can occur at only one altitude, that is, where $P_e = P_o$. Therefore, the altitude at which optimum expansion occurs depends upon the expansion ratio of the nozzle exit plane (A_e) divided by the area of the nozzle throat (A_t). The Greek letter, epsilon, is the designator for expansion ratio. Thus:

$$\epsilon = \frac{A_e}{A_t}. \tag{2}$$

Figure 3-3 shows the effect of nozzle expansion ratio on thrust. If we cut the nozzle off to the left of point B, the exit pressure is greater than ambient ($P_e > P_o$), and the nozzle underexpands (A_e/A_t is too small). The exhaust gases complete their expansion outside the nozzle.

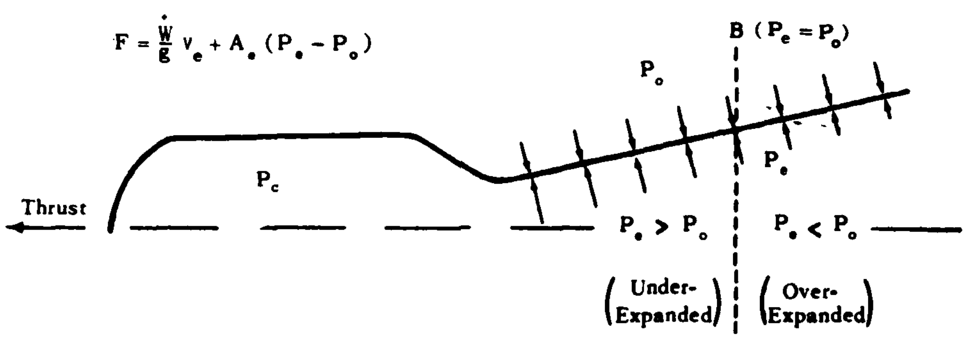


Figure 3-3. Nozzle expansion.

If we extend the nozzle beyond point B (further increasing the expansion ratio), the exhaust gas exit pressure will be less than ambient ($P_e < P_o$), and the nozzle will overexpand. Net thrust will be less because the pressure thrust loss, due to the increase in A_e, is greater than the momentum thrust gained from the increase in v_e.

Thus, for a given altitude (P_o) and a fixed nozzle, we obtain a maximum thrust when the nozzle has an expansion ratio so that $P_e = P_o$. As we fly this same engine to a higher altitude, the net thrust output will increase because of an increase in pressure thrust. Net thrust will not be as great as it could be if we could adjust the nozzle expansion ratio so that $P_e = P_o$ at all altitudes through which the engine operates.

In summary, the altitude where $P_e = P_o$ is the design altitude for a specific rocket engine. When $P_e = P_o$, the thrust output of the engine is the maximum that we can obtain at that altitude from that engine.

Engineers can design nozzles for optimum expansion at sea level or at any higher altitude. The designer selects the altitude for optimum expansion that gives the best average performance over the powered flight portion of the vehicle trajectory. In multistage rockets the expansion ratios vary in the vehicle stages that operate at the different altitudes. These multistage rockets use increasingly higher expansion ratios for those stages operating at the higher altitudes.

Altitude Effects and Thrust Parameters

In evaluating the preceding information, remember that thrust with a given nozzle (fixed area ratio), regardless of the design altitude at which optimum expansion occurs, increases with altitude due to decreasing P_o, until P_o is essentially zero. Large booster engines use nozzles that overexpand at launch, achieve optimum expansion at 40,000 to 50,000 feet and underexpand above 50,000 feet. In this way the nozzles provide higher thrust than if the engines expanded the nozzle optimally at launch. (See fig. 3-4).

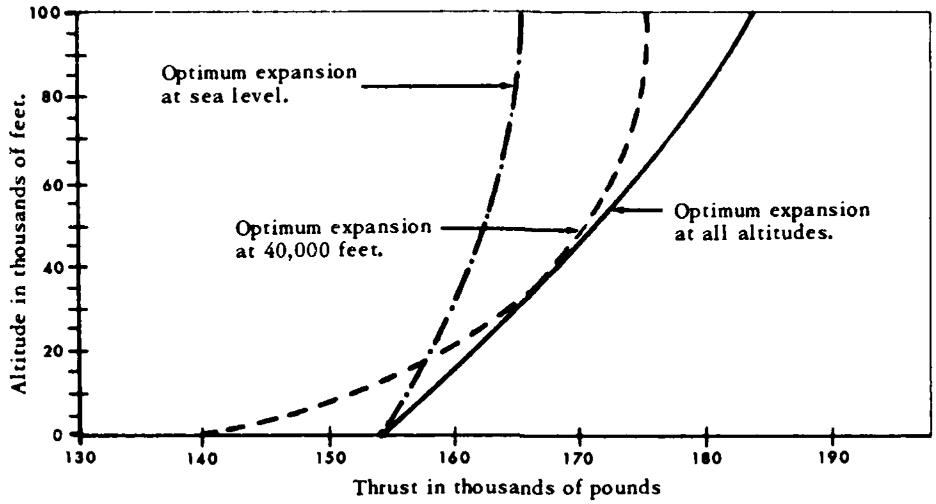


Figure 3-4. Thrust variation with altitude for nozzles of different expansion ratios.

Figure 3-5 shows which parameters increase as a specific rocket climbs to altitude, which parameters remain constant, and which parameters decrease in value. The rocket will reach an altitude where ambient pressure (P_o) becomes nearly zero. At this altitude, the thrust (F) and the specific impulse (I_{sp}) stabilize at a high value.

Mission Velocity Requirements

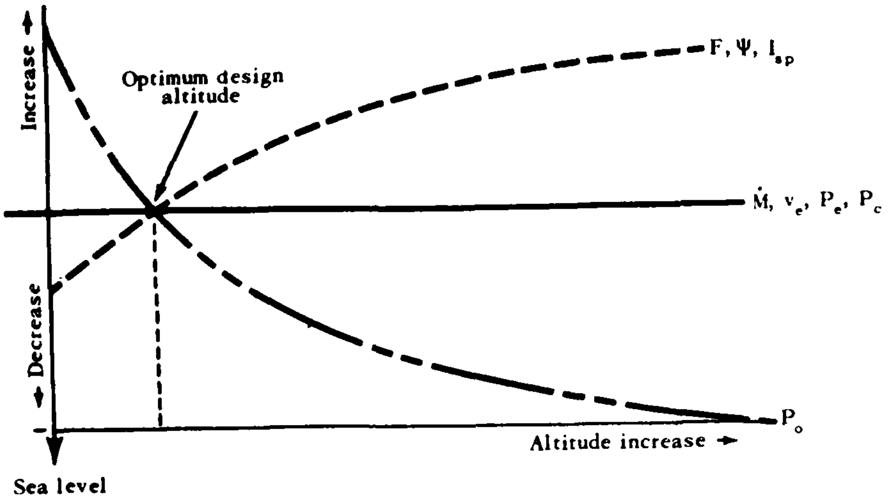
Figure 3-6 shows approximate values for the velocity and range relationships of a vehicle launched from the earth. Exact values depend on mission requirements, but the vehicle needs a specific velocity at the end of powered flight for each range—that is, Intermediate Range Ballistic Missile (IRBM) range—about 16,000 ft/sec; Intercontinental Ballistic Missile (ICBM)—24,000 ft/sec; and escape velocity near the earth's surface—36,700 ft/sec. Note that the velocity for ICBM range is very close to orbital velocities. In this area, small changes in velocity at the end of powered flight result in large changes in range.

Specific Impulse

Thrust is important in the definition of specific impulse:

$$I_{sp} = \frac{\text{Thrust (F)}}{\text{Weight rate flow of propellants } (\dot{W})} \qquad (3)$$

We express thrust (F) in pounds and propellant flow rate ($\dot{W}$) in pounds per second. Thus we express specific impulse (I_{sp}) in seconds. If the propulsion system has an I_{sp} of 300 seconds, it produces



Legend

F		Thrust	v_e	=	Velocity of gases at nozzle exit
Ψ	=	Thrust-to-weight ratio	P_e	=	Pressure of gases at nozzle exit
I_{sp}	=	Specific impulse	P_c	=	Combustion chamber pressure
$\dot{M}$	=	Propellant mass rate of flow $= \frac{\dot{W}}{g}$	P_o	=	Ambient atmospheric pressure

Figure 3-5. A summary of the common rocket engine thrust parameters versus altitude.

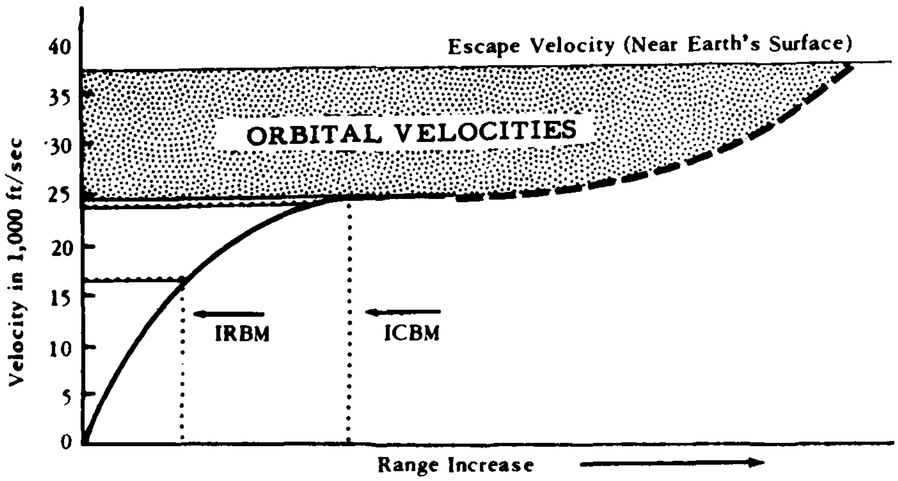


Figure 3-6. Velocity versus range of a rocket.

300 pounds of thrust for every pound of propellant burned per second. I_{sp} is one index of propulsive performance and relates to overall rocket performance.

Although this convention of expressing I_{sp} in seconds is almost universal, it is incorrect and misleading. Strictly speaking, specific pulse has the units:

$$I_{sp} = \frac{lb_f - sec}{lb_m}$$

The physical significance of I_{sp} becomes clear when we use metric units instead. Writing Newton's law, F = ma,

$$\text{Newton's (nt)} = \frac{kg - m}{sec^2}$$

$$I_{sp} = \frac{\text{Thrust}}{\text{Mass rate flow}} = \frac{nt}{kg/sec} = \frac{m}{sec}$$

Hence, I_{sp} is simply the gas velocity V_3 in equation 1.

Increasing I_{sp} improves the propulsion system's ability to increase vehicle velocity. This is the reason we quote the I_{sp} frequently to compare the performance of similar propulsion systems. We should not use I_{sp} alone to compare chemical with electrical or other dissimilar propulsion systems.

In designing a propulsion system, engineers must consider many compromises because of the many interrelated parameters. However, when they design a chemical propulsion system only to improve I_{sp}, there are two basic approaches. One is to design a better rocket engine. The other is to use better propellants. We discuss these two methods later in this chapter.

Mass Ratio

Mass ratio is a structural design parameter. It relates to propulsion because the difference between initial (or launch) and final (or empty) vehicle weights is the weight of propellant expended.

$$\text{Mass ratio} = \frac{\text{Initial Weight}}{\text{Final Weight}} = \frac{\text{Weight at Engine Start}}{\text{Weight at Engine Shutdown}} = \frac{W_1}{W_2} \qquad (4)$$

For example, a single-stage, World War II rocket had a mass ratio of 3 to 1 (expressed as $\frac{3}{1}$) and a range of approximately 200 miles. We computed the mass ratio for the rocket using these figures:

Item	*Launch Weight (lb)*	*Empty Weight (lb)*
Payload	2,000	2,000
Propellants	16,000	0
Other dry weight	6,000	6,000
Total	24,000	8,000

$$\text{Mass ratio} = \frac{\text{Launch Weight}}{\text{Empty Weight}} = \frac{24{,}000}{8{,}000} = \frac{3}{1}$$

It appeared that the way to propel a payload to greater ranges was to build bigger and better rockets. A bigger rocket could carry more propellants, permit longer thrust time, and achieve greater velocities and ranges. In improving the above rocket, the following ICBM design evolved:

Item	*Weight (lb)*
Payload	200 (note decrease in payload)
Propellants	1,254,000
Dry weight	29,800
Launch Weight	1,284,000

$$\text{Mass ratio} = \frac{W_1}{W_2} = \frac{(1{,}284{,}000 \text{ lb})}{(30{,}000 \text{ lb})} = \frac{42.8}{1}$$

Obviously, mass ratio had to increase markedly to achieve ICBM ranges. Single-stage rockets have not achieved high mass ratios. A good single stage rocket may have a $\frac{10}{1}$ mass ratio. The use of multistage rockets has increased mass ratios. The following ICBM, based on design criteria similar to, but slightly better than, the World War II rocket, shows the way that staging increases mass ratio.

	Item	*Weight (lb)*	
	Third Stage		
	Payload	200	= Vehicle weight at engine cutoff (W_2) (1,000 lb)
Stage weight	Dry weight	800	
	Propellant	2,500	
	Total	3,500	= Vehicle weight at engine start (W_1)
	Second stage		
	Payload	3,500	= Vehicle weight at engine cutoff (W_2) (10,000 lb)
Stage weight	Dry weight	6,500	
	Propellant	25,000	
	Total	35,000	= Vehicle weight at engine start (W_1)
	First stage		
	Payload	35,000	= Vehicle weight at engine cutoff (W_2) (65,000 lb)
Stage weight	Dry weight	30,000	
	Propellant	162,500	
	Total	227,500	= Vehicle weight at engine start (W_1)

This rocket carried the 200-pound payload to ICBM ranges just as the scaled-up rocket we previously described, but it had a gross weight at launch of only 227,500 pounds—about 17.7 percent of that of the scaled-up rocket. In the figures above, we use the initial weights (W_1) and final weights (W_2) for each stage to calculate the mass ratio for each stage operation. Remember, for each stage the initial weight is the weight of the vehicle when it starts the stage's engine, and the final weight is the weight of the vehicle when the vehicle shuts off the stage's engine. We must consider three mass ratios.

The third mass ratio is for the final stage only:

$$\text{Third-stage mass ratio} = \frac{W_1}{W_2} = \frac{3{,}500 \text{ lb}}{1{,}000 \text{ lb}} = \frac{3.5}{1}$$

The second mass ratio is for the remainder of the vehicle for second-stage operation:

$$\text{Second-stage mass ratio} = \frac{35{,}000 \text{ lb}}{10{,}000 \text{ lb}} = \frac{3.5}{1}$$

The first mass ratio is for the whole vehicle for first-stage operation:

$$\text{First-stage mass ratio} = \frac{227{,}500 \text{ lb}}{65{,}000 \text{ lb}} = \frac{3.5}{1}$$

The overall mass ratio of a multistage rocket is the product of the stage mass ratios. Therefore, in the three-stage rocket the overall mass ratio is:

$$\text{Overall mass ratio} = \left(\frac{3.5}{1}\right)\left(\frac{3.5}{1}\right)\left(\frac{3.5}{1}\right) = \frac{42.8}{1}$$

Staging reduces the launch size and weight of the vehicle required for a specific mission and aids in achieving the high velocities necessary for ICBM and space missions. The velocity of the multistage vehicle at the end of powered flight is the sum of velocity increases produced by each of the various stages. We add the increases because the upper stages start with velocities imparted to them by the lower stages.

Thrust-to-Weight Ratio

We express the comparison of engine thrust-to-vehicle weight as a thrust-to-weight ratio (Ψ).

$$\Psi = \frac{\text{Thrust (F)}}{\text{Weight of vehicle (W)}} \tag{5}$$

A vehicle launched vertically cannot lift off the surface of the earth unless Ψ is greater than 1 ($F > W$). The larger the value of Ψ, the higher the initial vehicle acceleration.

$$a = (\Psi - 1) \text{ g's} \tag{6}$$

The Ψ of the Minuteman missile is approximately 2, and its initial acceleration is about 1 g, compared with a Titan II missile with a Ψ of about 1.4 and an initial acceleration of about 0.4g.

IDEAL VEHICLE VELOCITY CHANGE

Specific impulse and mass ratio directly affect vehicle velocity. The vehicle performance equation shows how specific impulse and mass ratio relate to the magnitude of the ideal velocity change (Δv_i) of each stage at burnout or thrust termination.

$$\Delta v_i = I_{sp} \cdot g_c \cdot \ln \text{(MASS RATIO)} \tag{7}$$

The ideal Δv is the velocity change the rocket would attain if there were *no gravity, no drag,* and if the *earth did not rotate.*

Scientists derived this simplified equation from Newton's second law. Generally, they use it to find approximate Δv's. The equation will not provide the exact Δv, since atmospheric drag, the earth's rotation, and the changing effects of gravity will cause variations. A precise solution is a computer calculation, but this equation gives a reasonable approximation. Due to losses, the actual Δv's of earth-launched vehicles will be 4,000 to 6,000 ft/sec lower than those computed with this equation. They will vary with launch latitude and azimuth as well.

Equation 7 shows Δv to be directly proportional to two factors: I_{sp} and the natural logarithm (ln) of the mass ratio (MR). In other words, increasing either I_{sp} or MR, or both, can achieve higher Δv's. The equation does not consider physical size of the rocket or propulsion system. Therefore, in theory, different size rockets with identical I_{sp} and MR would achieve the same ideal Δv. Obviously, larger payloads require larger vehicles and higher thrust engines to achieve the same range or velocity.

The "g" in the equation comes from the conversion of mass to weight at the surface of the earth. Calculations in space use an altitude I_{sp}, which includes the increase in thrust with altitude but measures W at the surface of the earth. Therefore, g is *always* 32.2 ft/sec^2 in propulsion calculations.

Equation 7 applies to all rocket vehicles. For a one-stage rocket the Δv would be the final velocity at thrust termination with the mass ratio based on launch and burnout conditions. In multistage rockets, the ideal Δv is the sum of the Δv's developed by the various stages, with the I_{sp} and MR of each stage used to compute the Δv for that stage's burning time.

Engineers may use the equation to compute Δv for all space vehicles, including those with a stop and restart capability. When the vehicle applies power in flight, a mass ratio combining initial and final weights of the space vehicle for each powered phase is used in the calculations.

ACTUAL VEHICLE VELOCITY CHANGE

We can express actual launch vehicle velocity as:

$$\Delta v_a = \Delta v_i - \Delta v_l + v_r \tag{8}$$

In this equation:

Δv_a = Actual Δv

Δv_i = Ideal Δv (from equation 7)

Δv_l = Loss in Δv (gravity and drag)

v_r = Variation in Δv due to earth's rotation

Equation 7 may solve two types of problems. We have presented these problems on the succeeding pages. We can use math tables or the graph in figure 3-7 to determine natural logarithms. Note that, since the graph is actually a plot of natural logarithm (ln) values, we may extend it to accommodate values beyond the limits shown. We have plotted the graph so that mass ratios are on the vertical axis and the natural logarithms are on the horizontal axis. For example, if $(W_1/W_2) = 8$, we find this on the vertical axis. Then we find ln of 8 = 2.08 on the horizontal axis.

Now, we will discuss some sample rocket performance calculations. Scientists launched a rocket from the earth with an initial weight (W_1) of 296,000 pounds. At the end of powered flight, its final weight (W_2) was 40,000 pounds. Assuming no earth rotation and no velocity lost due to drag or gravity, calculate the magnitude of the ideal velocity (or Δv_i) at the end of powered flight. Assume a value of specific impulse

$$I_{sp} = 300 \text{ sec}$$

Use the equation: $\Delta v_i = I_{sp}\, g_c \ln \left(\frac{W_1}{W_2}\right)$

Math table method

a. Substitute

$$\Delta v_i = (300)(32.2) \ln \frac{296{,}000}{40{,}000}$$

$$\Delta v_i = (300)(32.2) \ln 7.4$$

b. Find: ln 7.4 in table.

c. Then:
$\Delta v_i = (300)(32.2)(2.0)$
$\Delta v_i = 19{,}320$ ft/sec

Graph method

a. Transpose equation to read:

$$\frac{\Delta v_i}{(I_{sp})(g_c)} = \ln \frac{W_1}{W_2}$$

b. Calculate value of W_1/W_2

$$\frac{296{,}000}{40{,}000} = 7.4$$

c. Use chart, figure 3-7.

d. Enter the graph where $W_1/W_2 = 7.4$ and then

e. Read the value of:

$$\frac{\Delta v_i}{(I_{sp})(g_c)} = 2.0$$

f. Then, solving for Δv_i

$\Delta v_i = (I_{sp})(g_c)(2.0)$

$\Delta v_i = (300)(32.2)(2.0)$

$\Delta v_i = 19{,}320$ ft/sec

This is a sample problem for changing an orbit. The following is a sample problem for finding the propellants to make a change in orbit.

A space vehicle is coasting in orbit. It has a restart capability and 3,500 pounds of propellants on board. We desire a change in orbit. Calculate the propellant required for the orbit change and whether we can complete the maneuver based on the following data.

Δv_a = 14,170 ft/sec = Magnitude of velocity change required to achieve the new orbit

I_{sp} = 400 sec = Specific impulse of the vehicle's engine at *altitude*

W_1 = 5,000 lb = Current weight of the vehicle in orbit (measured at the surface of the earth)

$g = 32.2 \frac{ft}{sec^2}$ = Acceleration due to gravity at earth's surface

1. Use the equation: $\Delta v_a = I_{sp}\, g_c \ln(W_1/W_2)$ (There are no drag or gravity losses in space so $\Delta v_i = \Delta v_a$)

2. Transpose to read: $\frac{\Delta v}{(I_{sp})(g_c)} = \ln\left(\frac{W_1}{W_2}\right)$

3. Solve for value of: $\Delta v/(I_{sp})(g_c)$

$$\frac{\Delta v}{(I_{sp})(g_c)} = \frac{14{,}170}{(400)(32.2)} = 1.1\text{; therefore } 1.1 = \ln(W_1/W_2)$$

4. Complete solution by either of following methods

Math table method

a. Since: $\ln(W_1/W_2) = 1.1$

b. Then: $W_1/W_2 = 3.0$

c. Then solving for W_2:

$$W_2 = \frac{W_1}{3.0} = \frac{5{,}000}{3.0} = 1{,}667 \text{ lb}$$

Graph method

a. Use graph, figure 3-7.

b. Enter the graph where: $\Delta v/(I_{sp})(g_c) = 1.1$ and then read the value of W_1/W_2, which is 3.0

c. Then solving for W_2:

$$W_2 = \frac{W_1}{3.0} = \frac{5{,}000}{3.0} = 1{,}667 \text{ lb}$$

d. For both above methods, let W_p = lb of propellants consumed;

e. Then: $W_p = W_1 - W_2 = 5{,}000 - 1{,}667 = 3{,}333$ lb of propellants consumed.

We can make the orbital change because the amount of propellants required is less than the 3,500 pounds that are available.

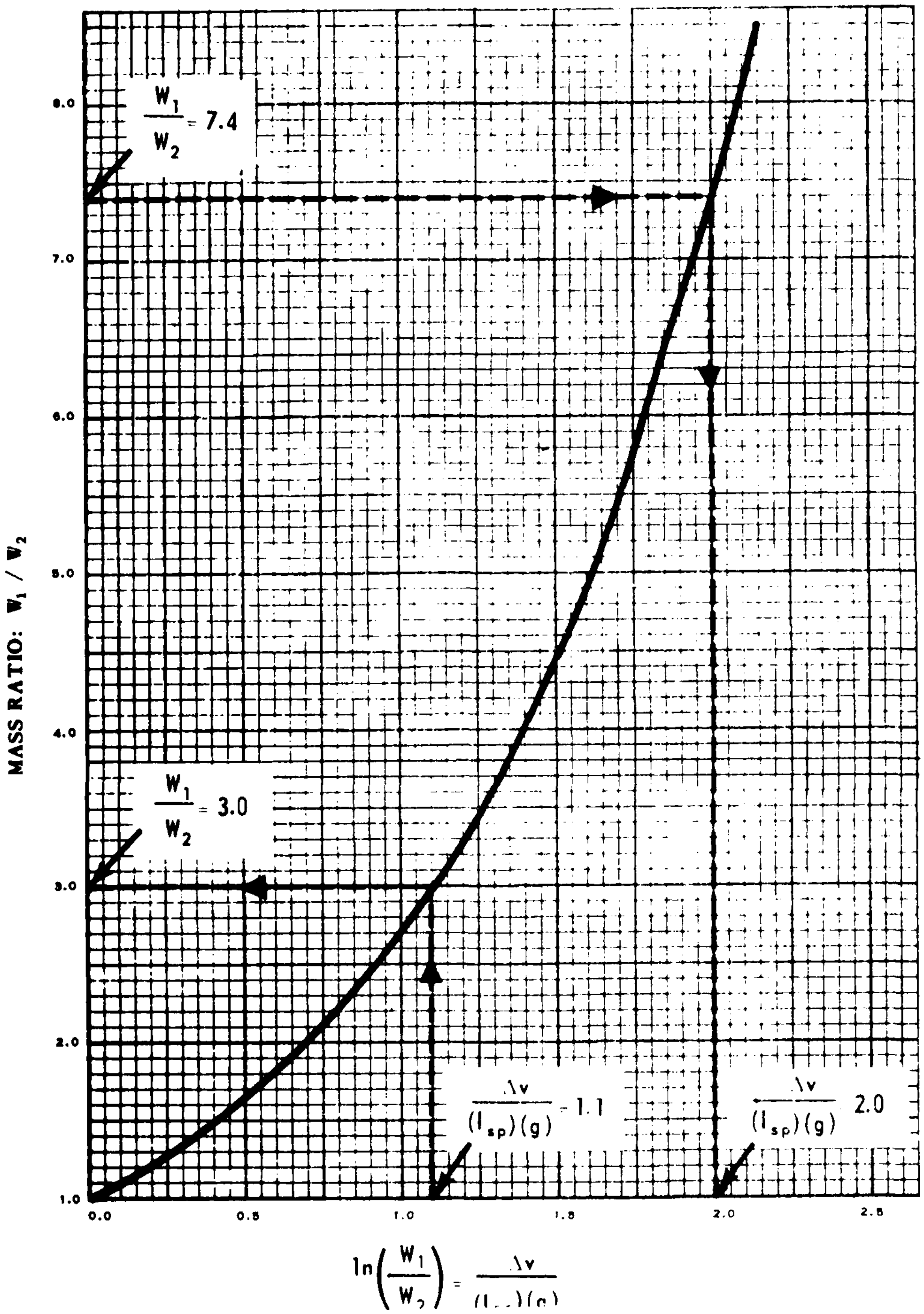


Figure 3-7. Mass ratio versus $\Delta v / (I_{sp})(g)$ or $\ln(W_1/W_2)$

PARKING ORBITS

Most missions that proceed beyond earth orbit begin by placing a payload into a parking orbit around the earth. The time in the parking orbit is used to wait for appropriate phase angles, for adjustments in launch windows, or for verifying equipment performance before proceeding on the next leg of the mission.

The optimum for a parking orbit from a drag and gravity standpoint lies somewhere between 85 and 100 nautical miles. The choice of 100 nautical miles allows optimum time in orbit for a manned vehicle.

Although the prediction of orbital lifetime is as yet an inexact science, these approximate values illustrate the choices available for parking orbits:

Altitude (nautical miles)	*Expected Time in Orbit (Days)*
85	½
100	3
150	35
200	200
300	4,000

ROCKET PROPELLANTS

Propellants are the working substances that rocket engines use to produce thrust. These substances may be liquid, solid, or gas, but the use of liquids or solids permits a particular rocket to carry more chemical energy. The space vehicle would require large, bulky tanks that contain only a small mass of compressed gas.

This section of the text considers only working substances that are accelerated by the energy released in their own chemical combustion (burning). The process involves a fuel and oxidizer reacting chemically to produce high-temperature, high-pressure gases. We call such fuels and oxidizers chemical propellants. Two chemical propellant subject areas are important: theoretical performance characteristics and energy content.

Theoretical Performance of Chemical Propellants

Consider a simplified picture of what happens in a rocket combustion chamber and nozzle (fig. 3-8). Burning the propellants releases large amounts of energy and produces high-temperature, high-pressure gases. The high temperature at point A in the combustion chamber is associated with very rapid motion of the gas molecules. These molecules have high speeds but move in random directions. As they approach the nozzle throat (B), their motion is less random, and they move toward the nozzle exit. At the nozzle exit (C), the largest velocity component of the molecules is parallel to the nozzle axis. The combustion chamber converts the chemical energy of the propellants to high-temperature random motion of the gas molecules, and the nozzle orients the velocity or kinetic energy (energy of motion) that gives the rocket its thrust.

As pointed out before, the velocity change of a space vehicle is a function of specific impulse (I_{sp}) and mass ratio. A higher I_{sp} increases the magnitude of the vehicle Δv for a given mass ratio or decreases the mass ratio necessary for a given vehicle Δv. Thus, I_{sp} is a measure of how well an engine converts chemical energy into velocity.

Since each mission in space is associated with a required Δv, I_{sp} determines if it is possible to perform a space mission with the mass ratios that are obtainable. For example, a mission to launch from the earth, land on the moon, and return to earth requires a series of Δv's totaling about 59,000 miles per hour. If the engines use propellants with low I_{sp} and can produce Δv's of only 55,000 miles per hour for a particular mass ratio or payload, the space vehicle cannot complete the mission. We need engines and propellants with higher I_{sp} for more difficult space missions.

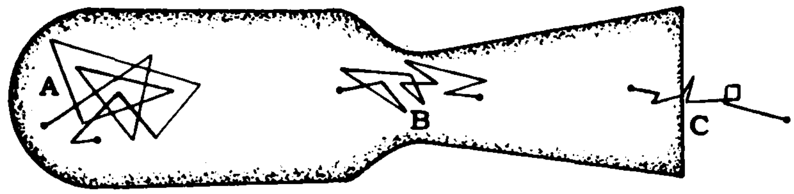


Figure 3-8. Velocity of gas molecules after combustion in a rocket engine.

Theoretical Specific Impulse

The theoretical specific impulse of chemical propellants in an ideal rocket is as follows:

$$I_{sp} = 9.797\sqrt{\left(\frac{k}{k-1}\right)\left(\frac{T_c}{m}\right)\left[1 - \left(\frac{P_e}{P_c}\right)^{\frac{k-1}{k}}\right]} \tag{9}$$

Where:

I_{sp} = Specific impulse (sec)
k = Average ratio of specific heats (C_p/C_v) of the combustion products. C_p is the specific heat of gases at constant pressure. C_v is the specific heat of gases at constant volume.
P_e = Nozzle-exit pressure (psia)
P_c = Combustion-chamber pressure (psia)
m = Average molecular weight of combustion products (lb/mole)
T_c = Combustion chamber temperature (°R). (Degrees Rankine is the absolute temperature and equals temperature in degrees F + 459.7°.)

The actual values of C_p, C_v, and k depend on the composition of the gas and its temperature. The combustion products in a rocket chamber and nozzle are a mixture of gases that vary in temperature from about 5,500° R in the combustion chamber to about 3,000° R at the nozzle exit; k is an average value of C_p/C_v for these temperatures (1.2–1.33 for chemical engines).

A molecule is the smallest quantity of matter that can exist by itself and retain all the properties of the original substance. For example, a molecule of water is the smallest quantity that has all the properties of water. A molecule of water contains two atoms of hydrogen whose combined atomic weight is about 2, and one atom of oxygen whose atomic weight is 16. Since a molecule is such a small amount of a substance, we use a mass numerically equal to the combined atomic weights, the molecular weight (called a mole), to give the equation a workable mass. In the English system of units a mole of water would be 18 pounds. We call it a pound-mole of water. Since the combustion products in a chemical rocket are a mixture of many gases, such as water vapor, carbon monoxide, carbon dioxide, hydrogen, and oxygen, we use the average molecular weight of the combustion products in the equation for theoretical specific impulse.

The combustion chamber temperature is the temperature obtained from the reaction of the oxidizer and fuel. This temperature depends on the mixture ratio, which is computed as follows:

$$\text{Mixture ratio (r)} = \frac{\text{Weight flow rate of oxidizer } (\dot{W}_o)}{\text{Weight flow rate of fuel } (\dot{W}_f)} \tag{10}$$

There are high and low limits to the possible mixture ratios for an oxidizer and fuel combination. If there is not enough oxidizer, or if there is too much oxidizer, combustion does not take place. A mixture ratio that produces complete combustion of both oxidizer and fuel gives the highest temperature. Too much oxidizer, even within the combustion mixture ratio limits, results in a lower temperature and excess oxidizer in the combustion products. Too little oxidizer results in a lower temperature and unburned fuel in the combustion products. Thus, mixture ratio, combustion

chamber temperature, and average molecular weight of the combustion products interrelate with one another.

Consider the relative effects of the ratio of specific heats, pressure, average molecular weight of combustion products, and combustion chamber temperature on theoretical I_{sp} for these conditions:

$k = 1.2$ $\quad$ $m = 20$ lb/mole $\quad$ $P_c = 1{,}000$ psia

$T_c = 6{,}000°R.$ $\quad$ $P_e = 14.7$ psia

$$I_{sp} = 9.797 \sqrt{\left(\frac{1.2}{0.2}\right)\left(\frac{6{,}000}{20}\right)\left[1 - \left(\frac{14.7}{1{,}000}\right)^{\frac{0.2}{1.2}}\right]}$$

$$I_{sp} = 9.797 \sqrt{(6)(300)(0.504)}$$

The ratio T_c/m (300) dominates the equation and:

$$I_{sp} \alpha \sqrt{\frac{T_c}{m}} \qquad (11)$$

This states that the theoretical I_{sp} of chemical propellants is directly proportional to the square root of the ratio of the combustion chamber temperature and the average molecular weight of the combustion products. The actual calculation of theoretical I_{sp} is complicated. The solution requires an extensive computation, which is best done with an electronic computer.

Figure 3-9 shows the relationship among specific impulse, combustion chamber temperature, and the molecular weight of the combustion products. Note that I_{sp} increases as the value of T_c/m increases.

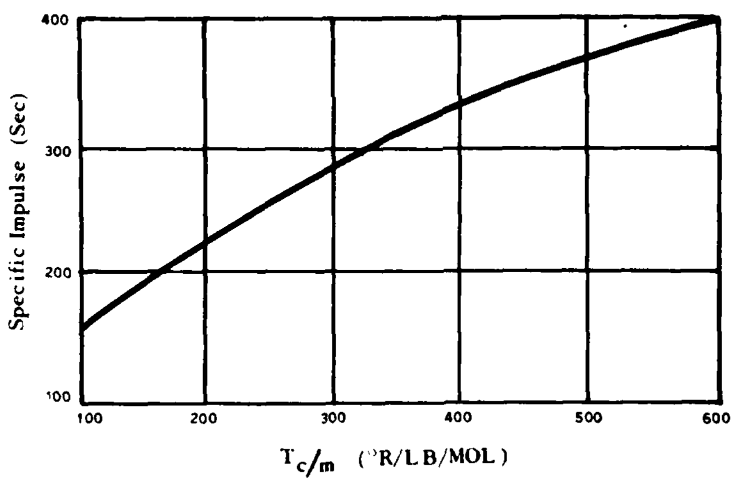


Figure 3-9. Specific impulse versus T_c/M.

The highest T_c may not produce the highest I_{sp}, since a corresponding increase in the average molecular weight of the combustion products may result in a smaller value for T_c/m. For example, most rockets using a hydrocarbon fuel like RP-1 (highly refined kerosene) burn a fuel-rich mixture that does not produce the highest T_c. This fuel-rich mixture produces low molecular weight gases like carbon monoxide and results in low average molecular weight combustion products, a high value for T_c/m, and a high I_{sp}. Burning a mixture of liquid oxygen (LOX) and RP-1, which produces the

highest T_c, results in high-molecular weight gases like carbon dioxide, a lower value for $T_c m$, and a lower I_{sp}. (See figs. 3-10a, 3-10b, 3-10c, 3-10d.)

The importance of low molecular weight combustion products effectively limits rocket fuels to chemical compounds of light elements, such as hydrogen, lithium, boron, carbon, and nitrogen. The use of elements heavier than aluminum (atomic weight 27) generally results in lower I_{sp}. Since hydrogen has the lowest molecular weight of all the elements, a propellant with a high hydrogen content is desirable.

In actual practice, the highest performance rocket fuels are relatively rich in hydrogen. Table 3-1 lists the *theoretical* I_{sp}s for various typical liquid propellant combinations. These are calculated for the following conditions: a combustion chamber pressure of 1,000 pounds per square inch of area, an optimum nozzle expansion ratio, an ambient pressure of 14.7 pounds per square inch of area, and a state of shifting equilibrium. The state of shifting equilibrium assumes a condition of changing chemical composition of gaseous products throughout expansion in the rocket nozzle. This changing chemical composition results in a higher I_{sp} than exists when the chemical composition of the gaseous products is fixed throughout expansion in the nozzle. We call the latter condition frozen equilibrium. The actual composition is somewhere between these two conditions.

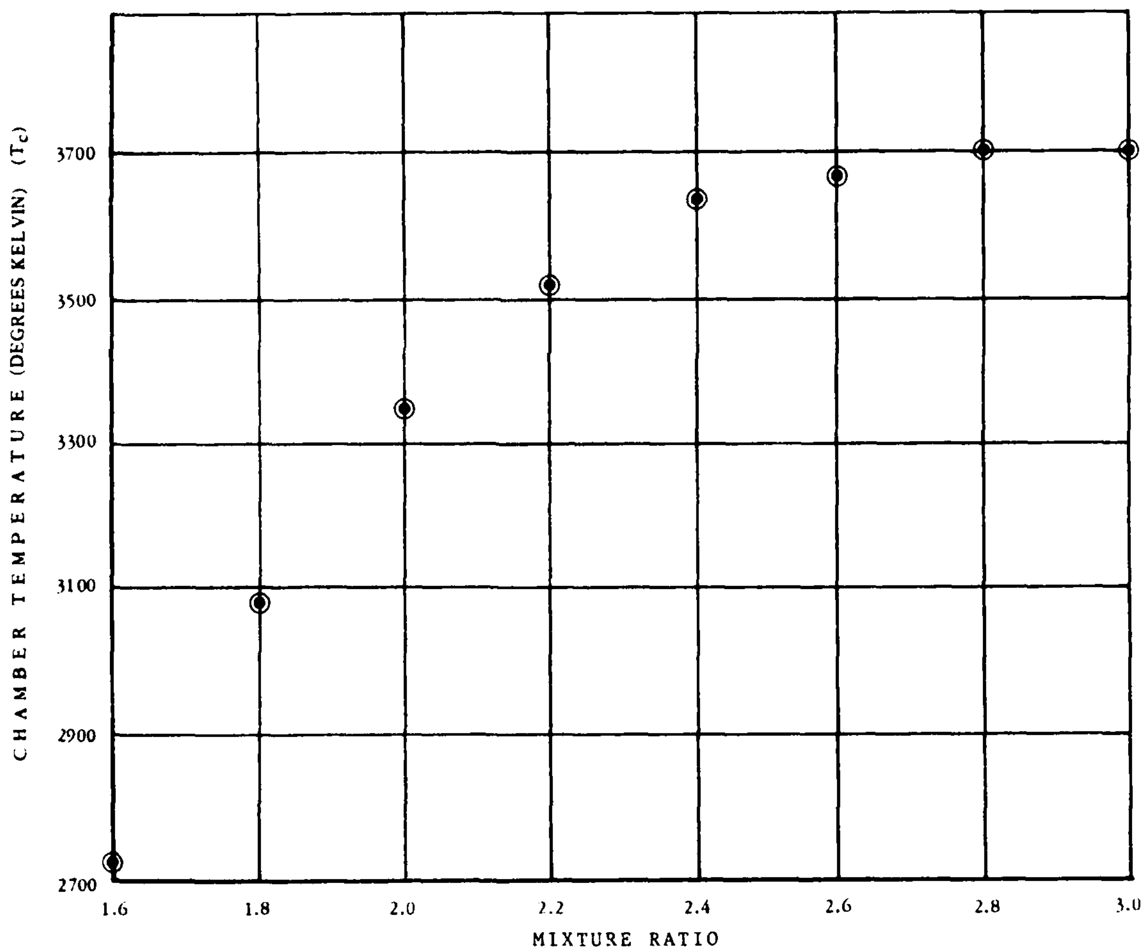


Figure 3-10a. Mixture ratio versus chamber temperature.

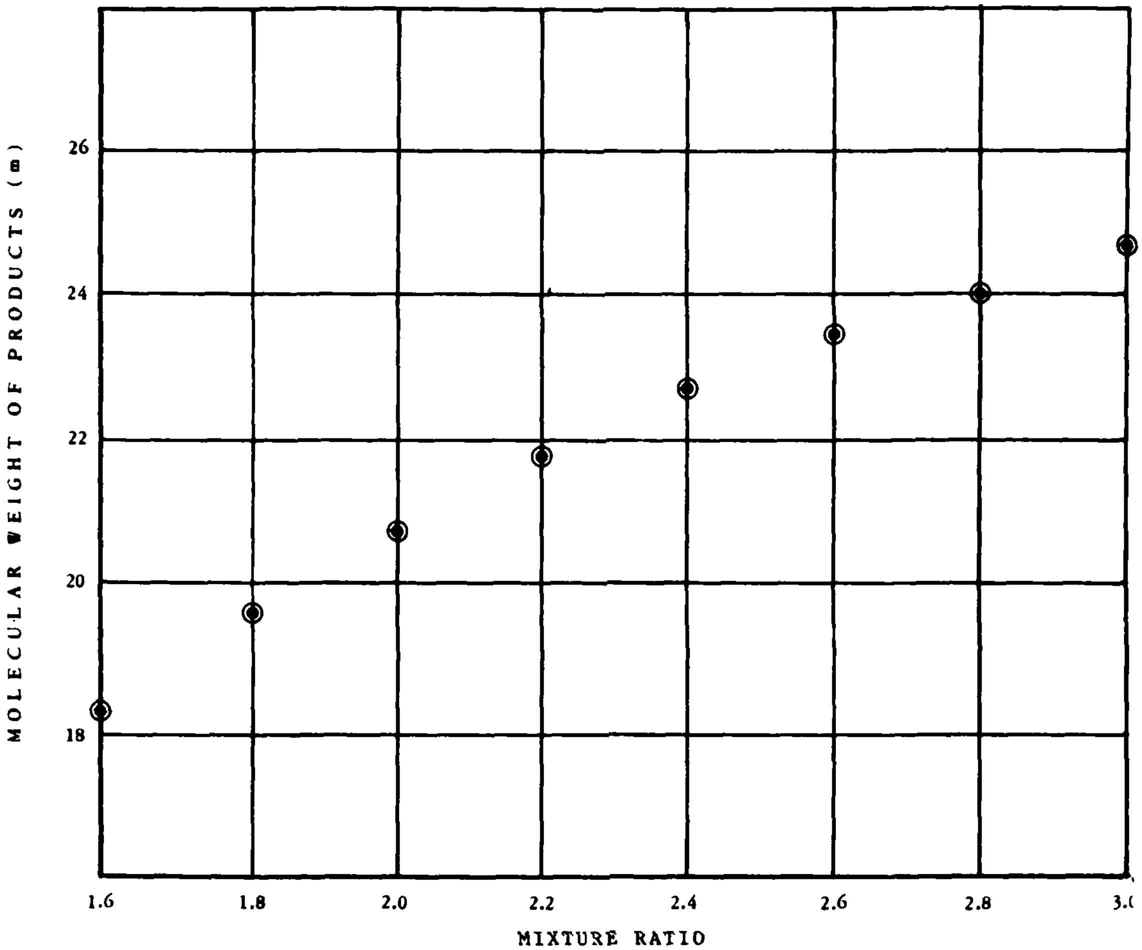


Figure 3-10b. Mixture ratio versus molecular weight of product.

Density Impulse

In some instances, a number of different propellant combinations may provide the same I_{sp}. Choosing a combination may then depend upon a parameter called density impulse. Density impulse is the product of specific impulse and specific gravity (SG) of the propellant:

$$I_d = (I_{sp})\ (SG) \qquad (12)$$

We use this parameter to relate propulsion system performance to the volume of the tank required to contain the propellants. Given a choice among propellant combinations each having the same I_{sp}, the combination with the highest SG requires smaller propellant tanks.

Total Impulse

Total impulse directly relates to the vehicle's ideal velocity change. Total impulse (I_t) relates the propulsion systems' thrust (F) to the time of operation or burning time (t_b). We define it as:

$$I_t = (F)\ (t_b) \qquad (13)$$

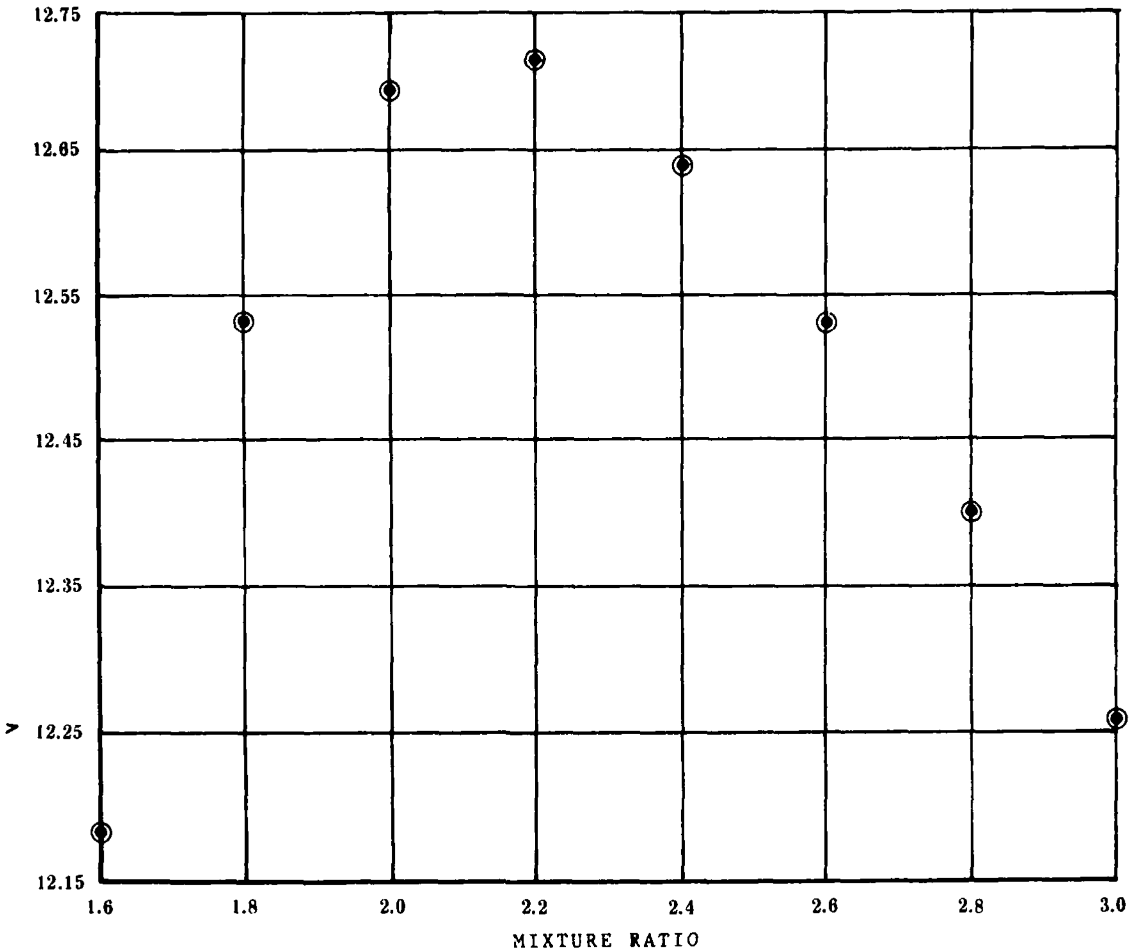


Figure 3-10c. Mixture ratio versus average molecular weight of combustion product.

Note that interchanging the values of thrust and operating time does not change the total impulse. For vertical flight the thrust must exceed the weight of the vehicle in order for it to accelerate. But, in orbit, either a high or low thrust would accelerate the vehicle to the same final velocity, provided the total impulse was the same. A higher thrust system accomplishes this through higher acceleration but for a shorter period of time than a low thrust system. Total impulse can be defined as:

$$I_t = (I_{sp})\ (W_p) \tag{14}$$

From this it is possible to see that we use this parameter to relate propulsion system performance to the allowable weight for propellants. Given a fixed volume and a choice among propellant combinations each having the same I_{sp}, the combination having the largest density would produce the greatest change in velocity.

Characteristics and Performance of Propellants

Rocket engines can operate on common fuels such as gasoline, alcohol, kerosene, asphalt or synthetic rubber, plus a suitable oxidizer. Engine designers consider fuel and oxidizer combinations

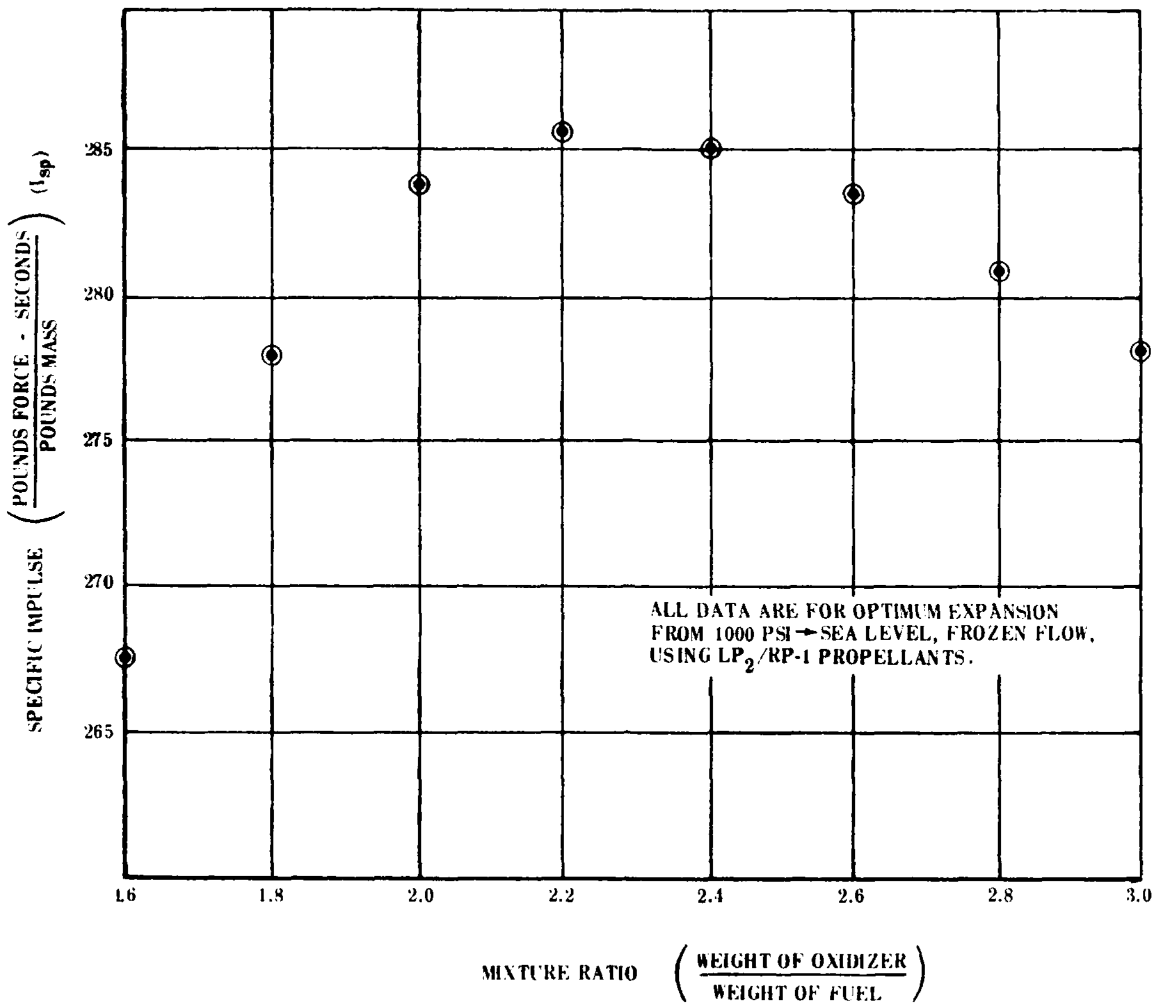


Figure 3-10d. Mixture ratio versus specifc impulse.

having the energy release and the physical and handling properties needed for desired performance. Selecting propellants for a given mission requires a complete analysis of mission; propellant performance, density, storability, toxicity, corrosiveness, availability and cost; size and structural weight of the vehicle; and payload weight.

Liquid propellants. The term "liquid propellant" refers to any of the liquid working fluids used in a rocket engine. Normally, they are an oxidizer and a fuel but may include catalysts or additives that improve burning or thrust. Generally, liquid propellants permit longer burning time than solid propellants. In some cases, they permit intermittent operations. That is, combustion can be stopped and started by controlling propellant flow.

Many combinations of liquid propellants have been investigated. However, no combination has all these desirable characteristics:

1. Large availability of raw materials and ease of manufacture.
2. High heat of combustion per unit of propellant mixture (for high I_{sp}).
3. Low freezing point (wide range of operation).

Table 3-1.
Specific Impulse of Liquid Propellant Combinations*

Oxidizer	FUEL					
	Ammonia	**RP-1**	**UDMH**	**50% UDMH and 50% hydrazine**	**Hydrazine (N_2H_4)**	**Hydrogen***
Liquid oxygen	294	300	310	312	313	391
Chlorine trifluoride	275	258	280	287	294	318
95% hydrogen peroxide and 5% water	262	273	278	279	282	314
Red fuming nitric acid (15% NO_2)	260	268	276	278	283	326
Nitrogen tetroxide	269	276	285	288	292	341
Liquid Fluorine	357	326	343		363	410

*Assumes P_c = 1000 psia, optimum nozzle expansion ratio and P_o = 14.7 psia.

4. High density (smaller tanks).
5. Low toxicity and corrosiveness (easier handling and storage).
6. Low vapor pressure, good chemical stability (simplified storage).

Liquid propellants can be classified as monopropellants, bipropellants, or tripropellants. A *monopropellant* contains a fuel and oxidizer combined in one substance. It may be a single chemical compound, such as nitromethane, or a mixture of several chemical compounds, such as hydrogen peroxide and alcohol. The compounds are stable at ordinary temperatures and pressures, but decompose when heated and pressurized, or when a catalyst starts the reaction. Monopropellant rockets are simple since they need only one propellant tank and the associated equipment.

The most common monopropellant system uses hydrazine. The development of a catalyst, Shell 405, that can spontaneously initiate decomposition at temperatures as low as the freezing point of the propellant made the system possible. Such systems have good I_{sp} values, long lifetimes, and good reliability. Space vehicles use these propellants for propulsion over a wide range of thrust values, and for generation of hot gases for such functions as propellant tank pressurization.

A *bipropellant* is a combination of fuel and oxidizer that are not mixed until after they have been injected into the combustion chamber. At present most liquid rockets use bipropellants. In addition to a fuel and oxidizer, a liquid bipropellant may include a catalyst to increase the speed of the reaction, or other additives to improve the physical, handling, or storage properties. Some bipropellants use a fuel and an oxidizer that do not require an external source of ignition, but ignite on contact with each other. We call these propellants hypergolic.

A *tripropellant* has three compounds. The third compound improves the specific impulse of the basic bipropellant by increasing the ratio T_c/m.

Liquid propellants are commonly classified as either cryogenic or storable propellants. A cryogenic propellant is one that has a very low boiling point and must be kept very cold. For example, liquid oxygen boils at −297° F, liquid flourine at −306° F, and liquid hydrogen at −423° F. Personnel at the launch site load these propellants into a rocket as near launch time as possible to reduce losses from vaporization and to minimize problems caused by their low temperatures.

A *storable propellant* is one that is liquid at normal temperatures and pressures and may be left in a rocket for days, months, or even years. For example, nitrogen tetroxide (N_2O_4) boils at 70°F,

unsymmetrical dimethyldrazine (UDMH) at 146° F, and hydrazine (N_2H_4) at 236° F. However, the term storable refers to storing propellants on earth. It does not consider the problems of storage in space.

Many rocket engines use one of the cryogenic propellants, liquid oxygen (Lo_2), with RP-1. The H-1 and F- engines of the NASA Saturn vehicles used this combination as do the engines of the Thor IRBM and Atlas ICBM missiles.

Liquid hydrogen (LH_2) and liquid oxygen (LO_2) comprise a cryogenic bipropellant. Upper stage engines, such as the RL-10 (Centaur engine) and the J-2 of the Saturn V, use this bipropellant.

The 50 percent UDMH-50 percent hydrazine fuel (Aerozine 50) with nitrogen tetroxide (N_2O_4) oxidizer is the storable hypergolic bipropellant used in the Titan III. It is classified as a bipropellant since the fuel contains two compounds to improve handling properties rather than improve I_{sp}.

Table 3-1 shows the improved performance of cryogenics. As a group, they have higher I_{sp} than the storable propellants.

The I_{sp} values in table 3-1 represent the maximum *theoretical* values for normal test conditions, which include engine operation at sea level. Actual engines using these propellants at sea level achieve 85 to 92 percent of these values. Engines operating near design altitude frequently achieve specific impulses that exceed these values.

For example, one version of the Atlas engine using RP-1 and LO_2 was designed for optimum expansion at 100,000 feet of altitude. The I_{sp} of this engine is 215 seconds at sea level and 309 seconds at 80,000 feet, compared with the 300 seconds listed in table 3-1. Scientists must analyze this I_{sp} reported for rocket engines, especially if it is much higher than the values in table 3-1. This analysis determines the altitude and other conditions for which the values are tabulated.

Several of the liquid propellants have theoretical I_{sp} approximately one-third higher than the conventional LO_2 and RP–1. LO_2 (liquid hydrogen) with LH_2 and LF_2 (liquid fluorine) with LH_2 are examples of bipropellants called *high-energy* propellants. The term "high energy" evolved from efforts to develop high-performance propellants. All the upper stages of large launch vehicles like the Saturn IB and Saturn V used these propellants.

Chamber pressure, altitude, and nozzle expansion ratio all affect engine performance, as we explained earlier. We can show clearly this effect on engine performance if you will consider figure 3-11. The dashed curves give I_{sp} values for a LO_2/LH_2 propellant combination at differing values of chamber pressure and expansion ratio. (The SSME curves are for the Space Shuttle main engines.) Notice first that a higher P_c gives higher I_{sp} for any value of ϵ. Second, there is an optimum ϵ for any P_c at sea level. This optimum value increases with increasing P_c. The solid curve, which is for expansion to vacuum, portrays the effect of altitude. Notice that this curve lies above the dashed curves for all ϵ and continues to increase to the right.

A comparison of the theoretical performance of the SSME and J-2 engines at increasing altitude also shows these effects. The higher P_c and ϵ of the SSME gives significantly better values of I_{sp} (fig. 3-12).

High-energy propellants in the upper stages of large rockets increase the payload or the mission capability of the vehicle. Consider two three-stage launch vehicles with the same initial gross weight (table 3-2). One has high-energy propellants in the upper stages; the other has conventional LO_2 and RP–1 in the upper stages. The one using high-energy propellants can carry a heavier payload for a given mission, or can perform a more difficult mission with the same payload. A three-stage vehicle with a gross weight of about one million pounds will have the theoretical payloads shown. For mission number 1, the vehicle using conventional propellants has less than half the payload of the one with high-energy propellants. For mission number 3, the conventional vehicle has only one-third the payload of the high-energy vehicle. Also, the conventional vehicle has the same payload (9 tons) for mission number 2 that the high-energy one has for mission number 3. Increasing the number of stages of the vehicle with LO_2h/RP–1 will increase the payload for the same initial gross weight, but will not approach that of the high-energy propellant vehicle.

Engineers design cryogenic fuel tanks differently from those for conventional fuel. Hydrogen tanks

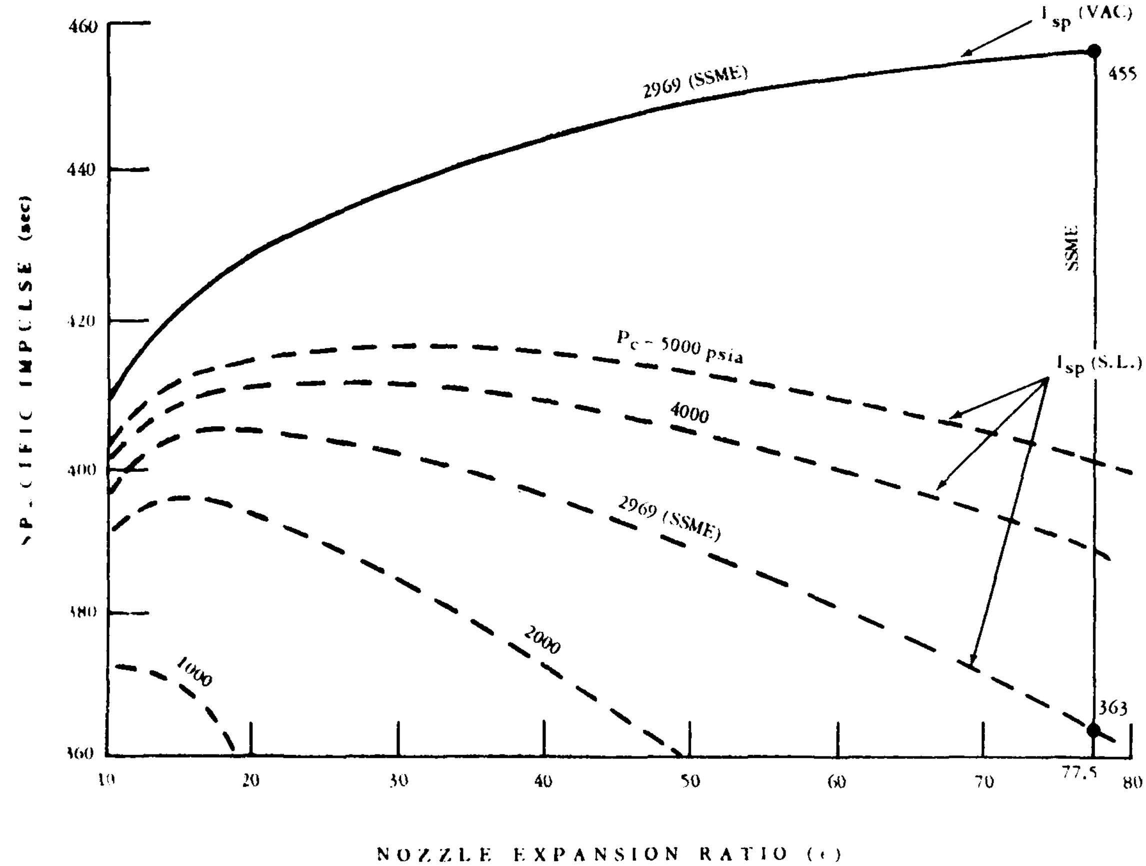


Figure 3-11. Nozzle expansion ratio (ε versus specific impulse (sec.).

are large and bulky because hydrogen has low density and low molecular weight. Therefore, upper stage weight and volume are larger if the upper stage uses hydrogen than if it uses conventional propellants.

Although high-energy propellants in upper stages increase mission capability, these propellants have high reactivity, and low temperatures. Because fluorine is very corrosive and toxic, it is difficult to handle and store. Fluorine, hydrogen, and oxygen are liquids only at very low temperatures. These low temperatures cause many metals to lose their ductility, and may cause the freezing of handling equipment such as valves. We find fluorine in relatively large quantities in nature but it is expensive to concentrate in a free state. Because of these and other problems, such as cost, high-energy propellants are usually used only in upper stages.

Solid propellants. Solid propellants burn on their exposed surfaces to produce hot gases. Solids contain all the substances needed to sustain combustion. Basically, they consist of either fuel and oxidizer that do not react below some minimum temperature, or compounds that combine fuel and oxidizer qualities (nitrocellulose or nitroglycerin). Scientists mix these materials to produce a solid with the desired chemical and physical characteristics.

We commonly divide solid propellants into two classes: composite (or heterogeneous), and homogeneous. Composites are heterogeneous mixtures of oxidizer and organic fuel binder. The fuel contains small particles of oxidizer dispersed throughout. We call this fuel a binder because the oxidizer has no mechanical strength. Usually a crystalline, finely ground oxidizer such as ammonium perchlorate is dispersed in an organic fuel such as asphalt. The oxidizer is approximately 70 to 80 percent of the total propellant weight. There are a large number of propellants of this type.

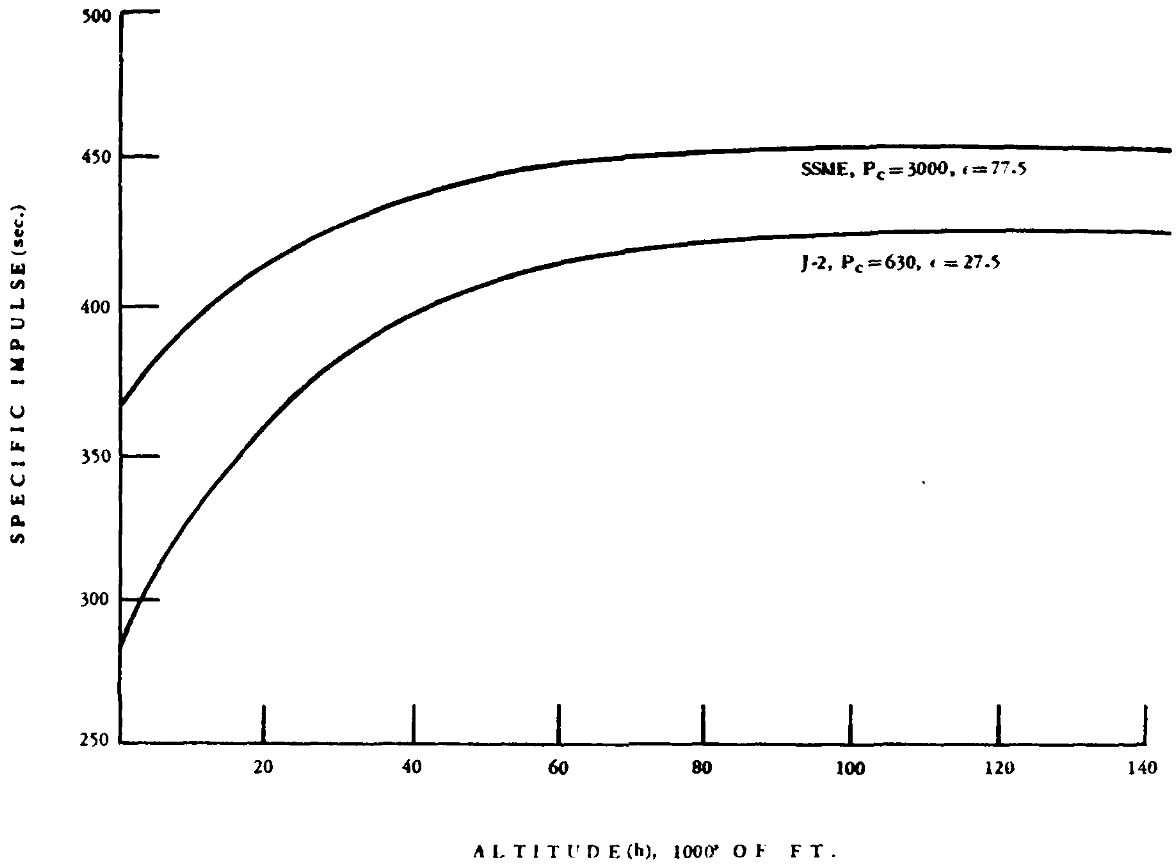


Figure 3-12. Effect of altitude on specific impulse.

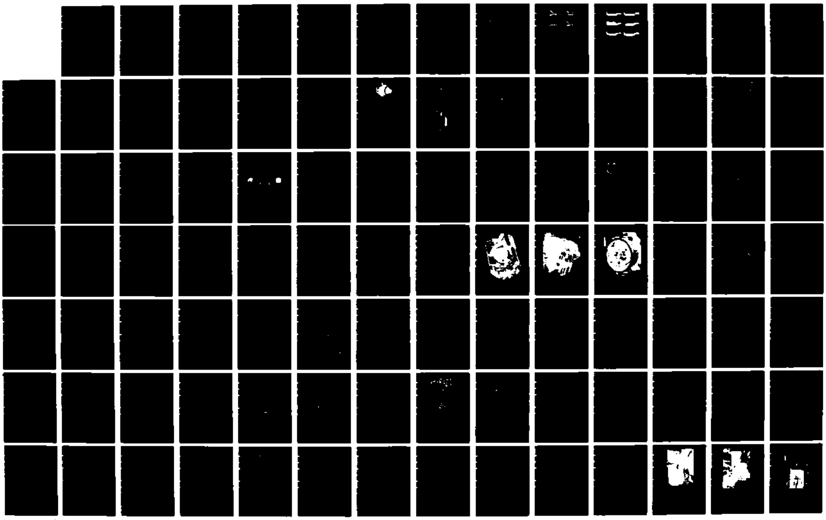
Table 3-2.
Comparison of Payloads for a Three-Stage Launch Vehicle Using Conventional Propellants With One Using High-Energy Propellants in the Upper Stages (Gross weight about one million pounds)

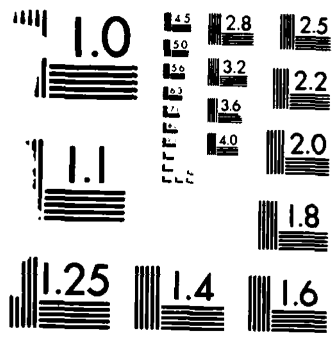
MISSION	PAYLOAD IN TONS	
	Conventional Upper Stages	*High-Energy Upper Stages*
Low earth orbit		
1. Ideal velocity ≈ 30,000 ft/sec	15	32
2. Ideal velocity ≈ 34,000 ft/sec	9	21
Escape		
3. Ideal velocity ≈ 41,000 ft/sec	3	9

Homogeneous propellants have oxidizer and the fuel in a single molecule. Scientists base most of these homogeneous propellants on a mixture of nitroglycerine and nitrocellulose, and call them "double-base propellants." The term distinguishes these propellants from many gunpowders that use either one or the other of the components as a base. Nitroglycerine is too sensitive to shock and has

AD-A165 038 SPACE HANDBOOK(U) AIR MINISTRY LONDON (ENGLAND) 2/4
C D COCHRAN ET AL. JAN 85 AU-18

UNCLASSIFIED F/G 15/7 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

too much energy to be used safely by itself in an engine. However, it forms a suitable propellant when combined with the less energetic but more stable nitrocellulose. The major components of a typical double-base propellant are:

Component	*Percent of total*
Nitrocellulose	51.38% (propellant)
Nitroglycerine	43.38% (propellant)
Diethyl phthalate	3.09% (plasticizer)
Potassium nitrate	1.45% (flash depressor)
Diphenylamine	0.07% (stabilizer)
Nigrosine dye	0.10% (opacifier)

Note the additives that control physical and chemical properties. Each additive performs a specific function. The *plasticizer* improves the propellant's structural properties. The *flash depressor* cools the exhaust gases before they escape to the atmosphere and promotes smooth burning at low temperatures. The *stabilizer* absorbs the gaseous products of slow decomposition and reduces the tendency of the propellant to absorb moisture during storage. The *opacifier* prevents heat transfer by radiation to sections of the propellant that have not started to burn. (Small flaws in the propellant can absorb enough heat through radiation to ignite the propellant internally, producing enough gas to break it up if an opacifier is not present.)

An ideal solid propellant would possess these characteristics:

1. High release of chemical energy.
2. Low molecular weight combustion products.
3. High density.
4. Readily manufactured from easily obtainable substances by simple processes.
5. Safe and easy to handle.
6. Insensitive to shock and temperature changes and no chemical or physical deterioration while in storage.
7. Ability to ignite and burn uniformly over a wide range of operating temperatures.
8. Nonhygroscopic (nonabsorbent of moisture).
9. Smokeless and flashless.

It is improbable that any propellant will have all of these characteristics. Propellants used today possess some of these characteristics at the expense of others, depending upon the application and the desired performance.

The finished propellant is a single mass called a *grain* or stick. A solid propellant rocket has one or more grains that constitute a charge in the same chamber. We found the use of solid propellants limited until the development of high-energy propellants and processing techniques for making large grains. Now we can make single grains in sizes up to 22 feet in diameter.

In addition to being composite or homogeneous, we class solid propellants as restricted or unrestricted. They are restricted burning charges when we use inhibitors to restrict burning on some surfaces of the propellant. Inhibitors are chemicals that do not burn or burn very slowly. Controlling the burning area in this manner lengthens the burning time and results in lower thrust. An inhibitor applied to the wall of the combustion chamber reduces heat transfer to the wall. We call this a liner or insulation.

Charges without an inhibitor are *unrestricted* burning charges. These burn on all exposed surfaces simultaneously. The unrestricted grain delivers a large thrust for a short time, whereas a restricted grain delivers smaller thrust for a longer time. Today, most large solid-propellant rockets contain restricted burning charges.

The operating pressure, thrust, and burning time of a solid-propellant rocket depend upon the chemical composition of the propellant; its initial grain temperature; the gas velocity next to the burning surface; and the size, burning surface, and geometrical shape of a grain. A given propellant can be cast into different grain shapes with different burning characteristics.

The thrust of a rocket is proportional to the product of the exhaust velocity and the propellant flow

rate. Large thrust requires a large flow rate that a large burning surface, a fast burning rate, or both, produces. The speed of the flame passing through a solid propellant in a direction perpendicular to the burning surface determines the burning rate of a solid propellant. Burning rate depends on the initial grain temperature, and upon the operating chamber pressure. A solid rocket operates at a higher chamber pressure and thrust when the propellant is hot than when it is cold, but it will burn for a shorter time. The converse is also true. If the chamber pressure is below a minimum value, the propellant will not burn. Variation in the geometric shape of the grain changes the burning area and thrust output. Figure 3-13 shows several typical grain shapes.

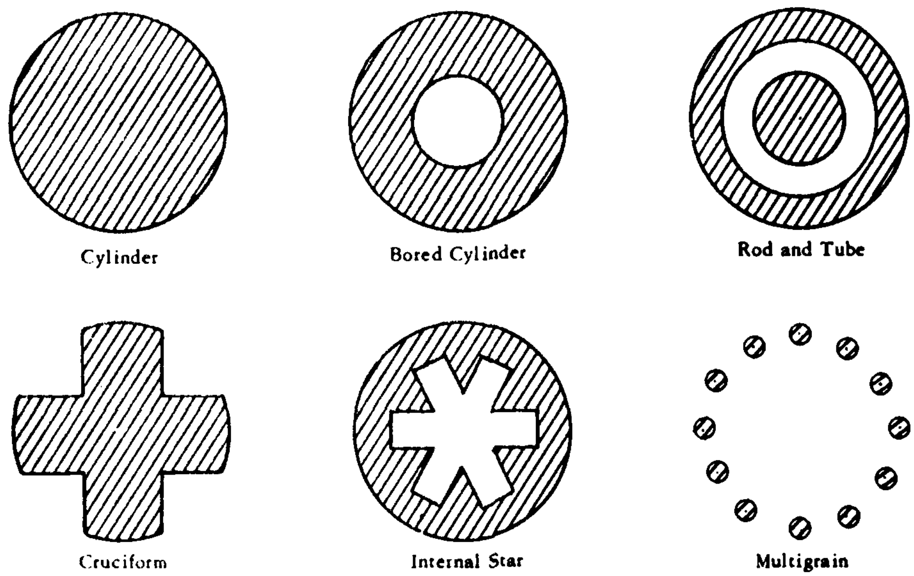


Figure 3-13. Typical solid propellant grain shapes shown in cross section.

Basically, there are three ways in which the burning area can change with time. If the area increases as burning progresses, the thrust increases with time. We call this grain a progressive burning grain. If the burning area decreases with time, thrust decreases. We call this grain a regressive burning grain. If the area remains approximately constant during burning, thrust is constant. We call the grain a neutral burning grain. In each case, the type of burning determines how the thrust level changes with time, following initial thrust buildup.

Regressive burning does not produce as much peak acceleration as does neutral or progressive burning, because the thrust decreases as vehicle weight decreases. Lower accelerations are required when rockets are used to propel payloads that cannot withstand high acceleration loading. Figure 3-14 shows progressive, neutral, and regressive burning.

The progressive grain increases burning surface area and proportionally the thrust as time increases. The bored cylinder inhibited at both ends and at the chamber wall is an example.

The neutral grain maintains a constant burning surface area and thrust over time. A rod and tube grain inhibited at the ends and at the chamber wall is an example.

The regressive grain's burning surface area decreases over time thus decreasing thrust over the same time. A cylinder with both ends inhibited is an example of a regressive grain. Table 3-3 shows the theoretical I_{sp} for typical solid propellants.

Some of the earliest composite propellants had an asphalt and fuel oil base with about 75 percent potassium perchlorate oxidizer. Scientists added oil to the asphalt to keep it from becoming brittle and cracking at low temperature, but the resultant propellant became soft and began to flow at high

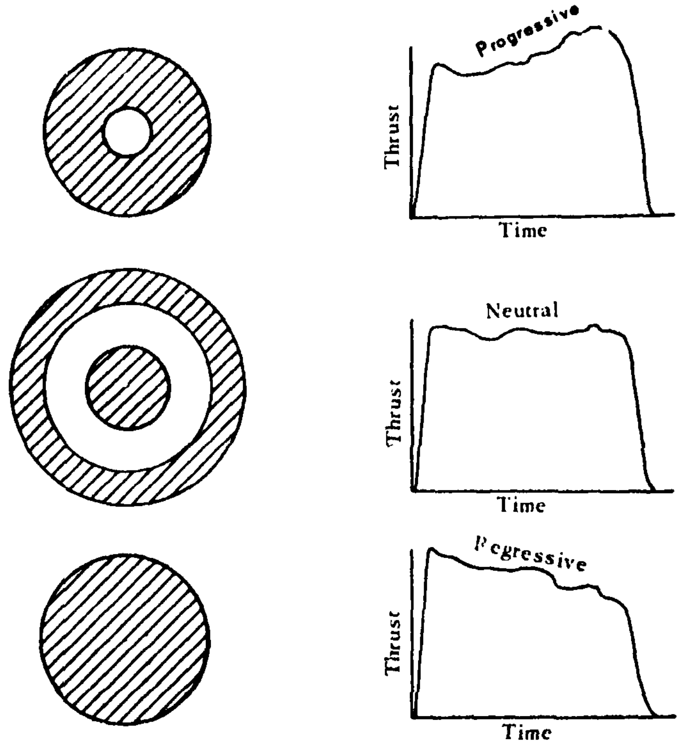


Figure 3-14. Progressive, neutral, and regressive burning grains.

Table 3-3.
Specific Impulse of Solid Propellant Combinations

Fuel base	*Oxidizer*	I_{sp} *(seconds)*
Asphalt	**Perchlorate**	**200**
Nitrocellulose and nitroglycerine	---	**240**
Polyurethane	**Perchlorate**	**245**
Carboxy-terminated Polybutadiene (CTPC)	Perchlorate	260
Hydroxy-terminated Polybutadiene (HTPB)	Perchlorate	260
Cross-linked Double Base	----	270
Boron	**Perchlorate**	**270**
Metallic hydride	**Fluoride**	**300**

temperatures. The specific impulse and operating range of these propellants was rather limited.

The double-base propellants use nitroglycerine and nitrocellulose with the proper additives. These propellents have a higher specific impulse than the asphalt-perchlorate propellants but are more sensitive to shock.

In the past the use of a higher percentage of oxidizer has improved the specific impulse of *composite propellants*. This was successful only as long as the propellant retained adequate structural properties, since the fuel binder gives the propellant its mechanical stength. The addition of a light metal, such as aluminum, to the fuel is another way to improve the specific impulse. Light metals increase the combustion chamber temperature and lower the molecular weight of the combustion products. The result is a higher value of T_c/m and specific impulse. However, metal in the exhaust gases may cause erosion of the rocket nozzle.

Replacing the inert fuel binder with a high-energy fuel binder containing oxygen (double-base propellants) further increases specific impulse. Some of the propellants used today combine a double-base propellant with a composite propellant and a metal fuel.

As a group, the solid propellants have a lower specific impulse than the liquid propellants, but they have the advantages of simplicity, instant readiness, low cost, potential for higher acceleration, and high density compared to the start-stop capability, high energy, regenerative cooling, and availability of the liquid propellants.

Hybrid propellants. We can combine some of the advantages of liquid and solid propellants in a hybrid rocket. In a hybrid engine, we store a liquid (usually oxidizer) in one container while we store a solid (usually fuel) in a second (see fig. 3-22). The separation of propellants in a hybrid eliminates the dependence of burning time on the grain temperature. The absence of oxidizer in the solid grain improves its structural proprties. The hybrid combines the start-stop advantages of liquid propellants with the high density, instant readiness, and potentially high acceleration of the solid propellants. It is a relatively simple system with high performance. Solid fuel lithium, suspended in a plastic base, burned with a mixture of fluorine and oxygen (FLOX) produces a theoretical I_{sp} of about 375 seconds.

Storage in space. Propellant storage in space is one of the problems that scientists must consider in selecting the chemical propellants for space propulsion. The system may use the propellants intermittently over long periods of time, or it may store them for months prior to peforming one maneuver. In space, the propellants no longer have the protection of the earth's atmosphere. They must perform in a hostile environment that includes a hard vacuum, thermal radiation from the sun, energetic particles such as cosmic rays, and meteoroid bombardment.

Measures are necessary to protect propellants from deterioration or loss by evaporation. It is important to protect liquid propellants from evaporation and to ensure that components are leakproof. Solid propellants must be protected from radiation that affects their burning rate and physical properties. Radiation may cause changes in liquid propellants. Propulsion systems must withstand temperature extremes when one side faces the sun and the other the coldness of space. They must also be protected from micrometeoroid damage. These are a few of the problems caused by the space environment. We need additional research and testing to define all the effects of space environment on propellants.

CHEMICAL ROCKET ENGINES

In the previous section we discussed in detail the basic method of improving I_{sp} by using more energetic propellants. Designing a better engine for a given propellant is the second basic method. Using only one method for improvement cannot solve the problem of achieving higher specific impulse. Frequently both methods are used to produce the best results.

This section describes a few typical examples of liquid propellant, solid propellant, and hybrid propellant engines, and suggests a few of the many possible design improvements. We have limited the material to the large and comparatively high-thrust rockets.

We may classify chemical rockets by several different methods, one of which refers to the type of propellants used. We should remember the groups shown in figure 3-15 when we make comparisons and discuss methods for improving performance, because the different groups have different characteristics.

In general, when we need fast reaction (as in military ICBMs), ease of handling and simplicity of operation are paramount. In such applications, solid propellant or storable liquid propellant engines are preferred.

For space missions requiring restarts, controllability, or high I_{sp}, liquid propellant engines are generally best. For any particular mission, there must be trade-offs made to determine the best engines suited for that mission. The hybrid engines are attempts to combine the advantages of both

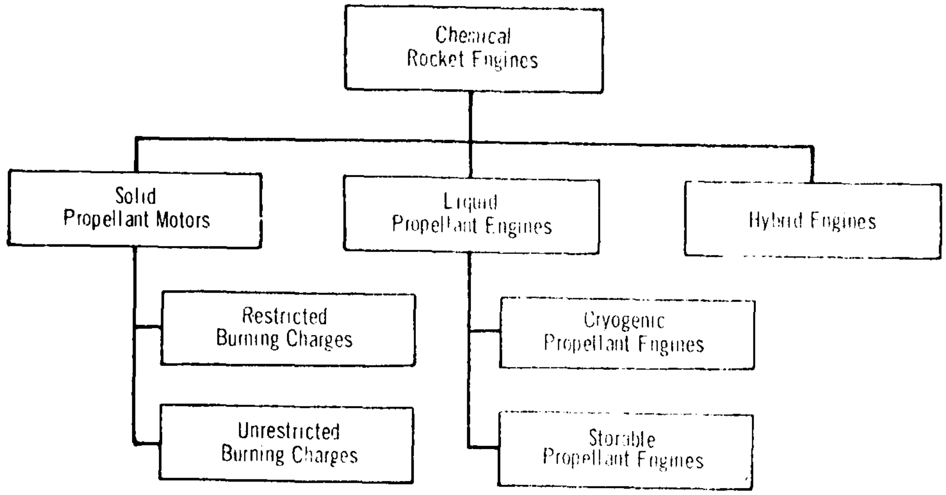


Figure 3-15. One method of classifying chemical propellant rockets.

liquid and solid engines. Several small hybrid engines are in use.

Rocket engines are very lightweight for the amount of thrust they produce. Generally, a rocket engine will deliver about 100 pounds of thrust for every pound of engine weight, compared to the better jet engines, which produce approximately 10 pounds of thrust per pound of engine weight. If the thrust of a vehicle's engines can be improved without increasing their weight, the propulsion system will be improved.

In comparing two or more engines to determine which is best for an application, designers must evaluate the entire vehicle and its mission. For example, consider two liquid-propellant engines with similar thrust levels. The engine with higher I_{sp} would be more efficient and have higher vehicle velocity potential. When we compare engines with similar operating characteristics, propellants, and thrust levels, higher I_{sp} results from better propellant utilization.

We might make a comparison between two similar liquid engines with similar propellant consumption rates, but different thrust levels. The engine with higher thrust yields higher I_{sp} and is more efficient. Designers try to achieve the highest possible I_{sp} for any given rocket because more I_{sp} yields increased velocity and payload.

Liquid-Propellant Engines

A liquid propulsion system consists of propellant tanks, propellant feed system, thrust chamber, and controls such as regulators, valves, and sequencing and sensing equipment. The propellants can be monopropellants, bipropellants, or tripropellants, and may be either storable or cryogenic fluids.

The least complex of these is one designed for monopropellants. Here, there is only one propellant tank, a single feed system (usually pressure-fed), and a comparatively simple injector (since the engine does not require a mixing fuel and oxidizer). Monopropellant rockets are in use today but do not yet develop high thrust. However, the simplicity of monopropellant engines makes them adaptable and frequently desirable for use in attitude control or small velocity corrections in deep space.

The liquid bipropellant system in common use is more complex. Figure 3-16 shows the basic components schematically. Note that this system requires two tanks, two feed systems, and multiple injectors.

Some bipropellant systems use pressure-fed propellant flow. Here, pressurizing the tanks with an inactive gas such as nitrogen or helium forces propellants from the tanks to the engine. Igniting either a solid-propellant grain or some of the vehicle's propellants in a gas generator designed for this

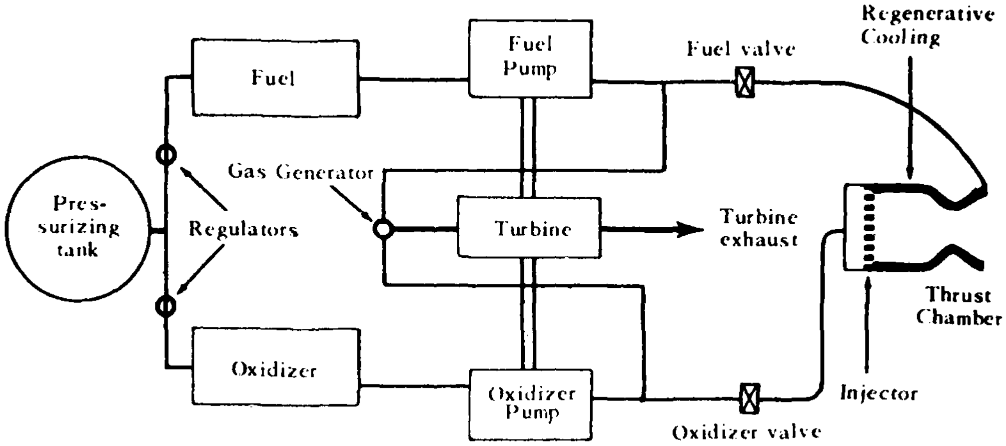


Figure 3-16. Schematic reproduction of a liquid bipropellant system.

purpose can create a pressurizing gas.

Pumps and gravity can feed bipropellants to the engine, as some of the earlier booster engines did. However, the most commonly used method today is a combination of pumps and pressurizing tanks to provide positive pressure to the pumps feeding the engines. The bipropellant system is complex in comparison to solid-propellant systems because of the multiple pumps, the need to maintain the correct oxidizer-fuel mixture, the effect of the injector design upon stable combustion, and the need for thrust chamber cooling. Nevertheless, these complexities are within the state of the art and most space missions use bipropellant systems.

Often the vehicle's engine uses a centrifugal pump driven by a gas turbine to pump the propellants through the injector into the combustion chamber. A separate gas generator supplies gases to drive the turbine, or the engine bleeds these gases from the combustion chamber. The development of reliable turbopumps presented many challenges in design, materials, testing, and operational use. In some instances, more than 50 percent of the design effort for an engine is devoted to the turbopump.

Turbopumps must pump fuel and oxidizer simultaneously at different rates and be able to withstand high thermal stresses induced by the 1,500° F turbine gases while pumping cryogenic fluids with temperatures as low as -423° F (liquid hydrogen).

Scientists must ensure that the vehicle can maintain adequate seals at these temperature extremes, since the propellants would explode if they were to come in contact inside the pump. Pump design is critical because the pumps must develop high propellant pressure. Higher combustion chamber pressure means higher I_{sp}, and pump outlet pressure must be higher than chamber pressure if propellants are to flow into the chamber.

Since scientists can design the components and controls of the liquid engine for individual control, the potentials of throttling and multiple restart make these engines attractive for in-space maneuvering.

Solid-Propellant Rocket Motors

People have used solid-propellant rocket motors for thousands of years. History tells of the ancient Chinese using skyrockets for celebrations as well as for weapons. The "rockets' red glare," as used in the national anthem, indicates the use of rockets during the War of 1812. Jet-assisted takeoff (JATO) units, to decrease aircraft takeoff roll or as takeoff assist units for lifting heavy loads, are familiar to most Air Force personnel.

The solid-propellant rocket is comparatively simple. The major components are: the case, which holds the propellants and is the combustion chamber; the igniter; and the converging-diverging nozzle (fig. 3-17). Because of its simplicity, the solid motor is inherently more reliable and cheaper to produce than the liquid rocket engine. However, solid motors generally have a lower I_{sp} than the liquid engines.

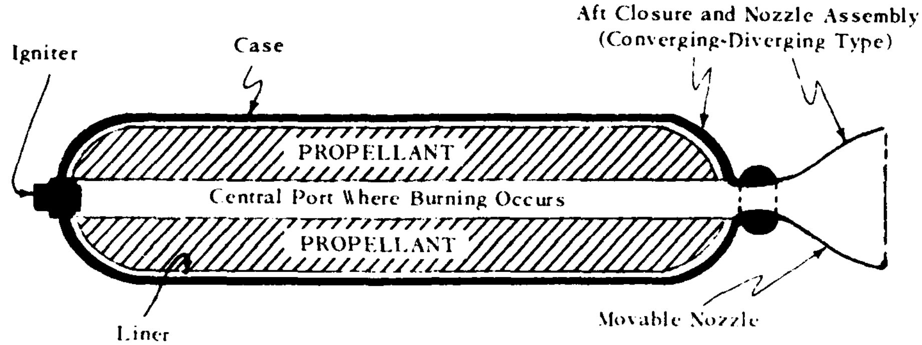


Figure 3-17. A solid propellant rocket motor

As mentioned earlier in the propellant section, the use of additives, new chemicals, and design of high volumetric loading propellant grains is improving the I_{sp} of solids. The other approach for increasing performance is to increase the mass ratio of the motor. Scientists and engineers have expended much energy in this area by designing cases that are lightweight and stronger. Research seems to point to two possible solutions: either make lightweight but strong cases of metals, such as titanium, or design filament-wound cases of fiberglass or nylon tape impregnated with epoxy-type glues.

We can make the filament-wound cases even lighter if we design reinforced propellant grains to assist in supporting the vehicle. We can form these reinforced grains by molding the propellant around aluminum or other metal additive wires. The engine consumes these reinforcing materials during combustion. Reinforced grain motors are usually regressive burning so that combustion chamber pressure will decrease near the end of burning to allow the use of very lightweight cases. Today industry can make both liquid and solid motors in a variety of sizes from very small attitude control and docking motors up to millions of pounds of thrust.

Thrust Vector Control

To accomplish attitude and directional control of a rocket, moving the engines or deflecting the exhaust gases is necessary. We call this thrust vector control. The use of flexible mountings or gimbals for the engine controls the thrust vector for the main and vernier liquid engines. Hydraulic or pneumatic cylinders deflect the engines to achieve the desired thrust vector.

In solid motors, deflecting or swiveling the entire motor is impractical because this amounts to moving the entire stage. Engineers must control thrust vectors at the nozzles in solid motors. They have used the following methods: movable nozzles; controllable vanes in the nozzles; jetavators (slip ring or collar at the nozzle exit); or injection of fluid (gas or liquid) into nozzles to deflect the exhaust flow and accomplish flight path or altitide changes (see fig. 3-18). These are only a few of the methods used.

Thrust Termination

Stopping the flow of propellants is all that is necessary to accomplish thrust termination. Since scientists can cause this to be done in a reproducible sequence, the residual thrust generated during,

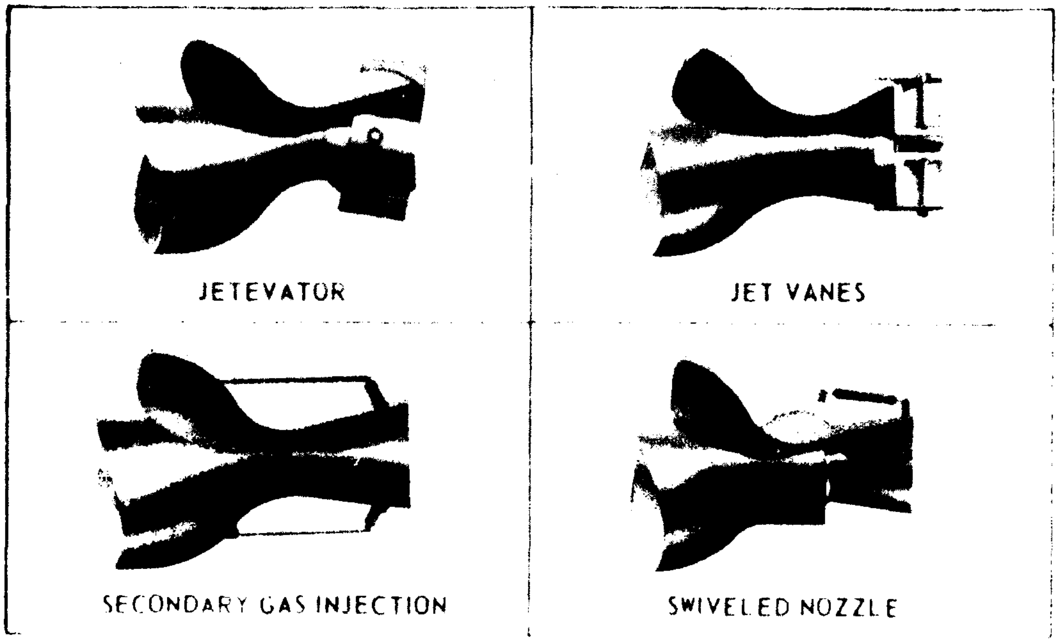


Figure 3-18. Methods of thrust vector control (TVC) for solid motors.

and shortly after, engine cutoff will be a known quantity. They can design liquid engines so that a rate variance in propellant flow into the engine can change the thrust level within limits.

In solid motors thrust termination is also simple (fig. 3-19). If we reduce the combustion chamber pressure below the critical pressure value for that particular motor, it will blow itself out, in effect. Blowing of nozzles, blowing out the aft end of the motor, or using thrust termination ports to vent the pressure out the sides of the case are some methods used to reduce the chamber pressure and terminate thrust. However, the thrust termination is not instantaneous. Low levels of thrust may continue for several minutes because of residual burning in the remaining solid grain. We must consider the residual burning characteristics of each type of motor when we need a very accurate cutoff velocity.

Engine Cooling

The combustion chamber of a rocket engine contains gases at temperatures in excess of 5,000° F. The walls of the combustion chamber and the nozzle absorb large amounts of heat. Some provision to dissipate heat must be made; otherwise, as the walls and nozzle absorb the heat, the wall temperature of the engine increases, and the strength of the structural material decreases. Since the amount of heat absorbed is highest in the nozzle throat, special attention is given to this region.

Most liquid engines are either partially or completely cooled. We cannot fire uncooled liquid rockets for periods greater than approximately 25 seconds. If combustion temperatures were low, it might be possible to use an uncooled engine for longer periods of time; but usually such is not the case in large engines. Cooling methods include regenerative, water, film, sweat (transpiration), radiation, and ablative cooling.

Regenerative cooling circulates fuel or oxidizer through small passageways between the inner and outer walls of the combustion chamber, throat, and nozzle. Transferring cools the engine and increases the energy of the propellant before it is injected into the combustion chamber. The addition

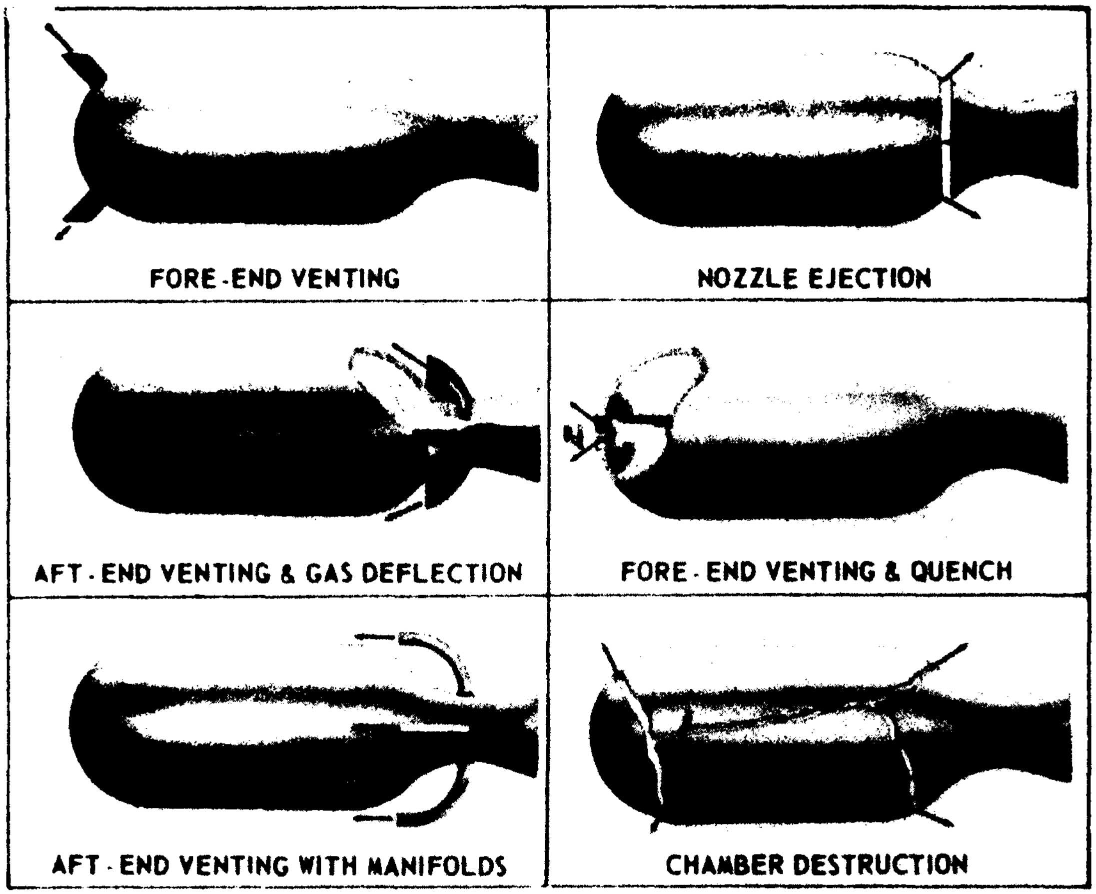


Figure 3-19. Methods of thrust termination of solid motors.

of energy to the propellant slightly increases the velocity of the exhaust gases and improves engine performance.

Water cooling is regenerative cooling, except that water circulates instead of the fuel oxidizer. Water cooling is widely used in static test stand firings, but not in flight vehicles.

Film cooling provides a thin fluid film to cover and protect the inner wall of the engine. The injection of small quantities of the fuel, oxidizer, or a nonreactive fluid at a number of points along the hot surface forms a protective film on the inner walls. The fluid flows along the wall and absorbs heat by evaporation. The engine can use film cooling with regenerative cooling for critical parts of the engine where regenerative cooling alone is not sufficient.

Sweat or transpiration cooling uses a porous material for the inner wall of the engine. The coolant is passed through this porous wall and is distributed over the hot surface. It is difficult to distribute the coolant uniformly over the surface because the combustion gas pressure decreases between the combustion chamber and the nozzle exit. Manufacturing porous materials so that they are uniform throughout is difficult, but is being done successfully, and transpiration cooling is being used in some applications.

Radiation cooling removes heat from the engine and radiates it to space. Some current liquid engines use a regeneratively-cooled nozzle to the 10:1 expansion ratio point and a radiation-cooled extension from that point to the end of the nozzle. This type is lighter than it would be if regenerative

cooling were used for the entire nozzle. Current research includes investigating radiation-cooled combustion chambers, nozzles, and nozzle extensions.

Ablative cooling involves coating the surface to be cooled with a layer of plastics and resins. The coating absorbs heat, chars, and then flakes off, carrying heat away from the subsurface being protected.

Liquid engines most commonly use the following cooling methods: regeneration cooling, film cooling, or a combination of the two, and ablative protection for some space engines. Since cooling a solid motor would require additional cooling equipment and materials, other protective means have been devised. Inserts of high-pressure metal alloys, graphite, or ceramics have been used with success. Using multiple nozzles and ablative materials in the nozzle keeps temperatures within working limits. Also, grain design and use of liners or inhibitors in the case control the heating. Combustion gas temperatures in solid motors are generally slightly lower than those in liquid engines.

Nozzles

The development of lighter weight, shorter nozzles has been in progress for several years. The reasons for this are (1) to reduce dead weight, (2) to reduce the interstage structure of multistage vehicles, and (3) to optimize the expansion ratio at all altitudes. Figure 3-20 shows several proposed improved nozzles.

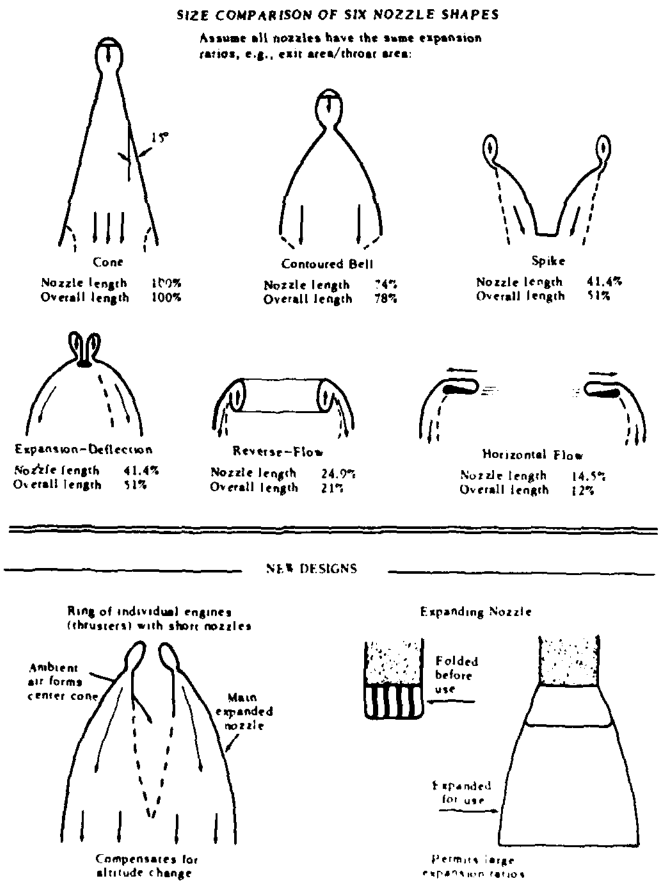


Figure 3-20. Nozzle design schematics.

Improvements

Several recent concepts have been used to develop bigger and better chemical rockets. Two such developments are the segmented solid and the hybrid rocket. Solid motors can be segmented or monolithic. The segmented motor is made by stacking separate grains (or segments) together to give the desired thrust level. Using segments greatly eases casting, inspection, and transportation problems for large motors. Special methods of assembly are made to prevent burning between segments, which would cause the motor to explode.

The solid segment idea (fig. 3-21) has been used to build large diameter motors. They manufacture these in four basic units or building blocks: the front closure segment, the center segment, the aft closure segment, and nozzle assembly. Varying the number of center segments can change the average thrust level of the rocket, thus controlling the burning area. For example, the Titan IIIC uses five-segment, 120-inch diameter, solid motors as boosters with each motor developing 1.2 million pounds of thrust. A seven-segment motor of this same diameter develops 1.6 million pounds of thrust. Above 156-inch diameter, the motors are so large that probably the engineer will have them manufactured or cast at the launch site, or transported by water. A monolithic (one piece) motor would be the simplest construction in this case.

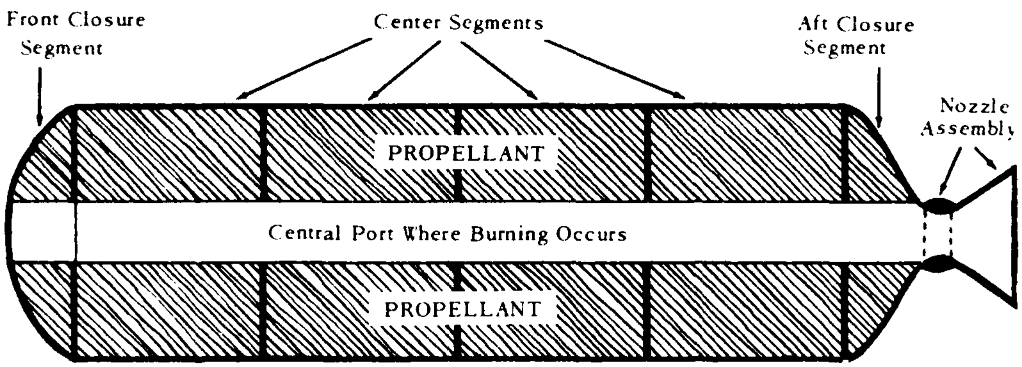


Figure 3-21. A segmented solid propellant rocket motor.

In the hybrid engine (fig. 3-22), liquid and solid propellants are used in one engine. The hybrid represents a compromise. It attempts to take advantage of the simplicity and reliability of solids; the higher performance and trottleability of liquids; storability and quick reaction time of solids; and, the hybrid reduces the hazard of having both fuel and oxidizer mixed together in one case.

In summary, present trends in liquid rocket engines include very high chamber pressures, high-energy propellants, and simplification of hardware. The use of lighter weight cases, thrust vector control, and better propellant formulations are improving solid-state rockets.

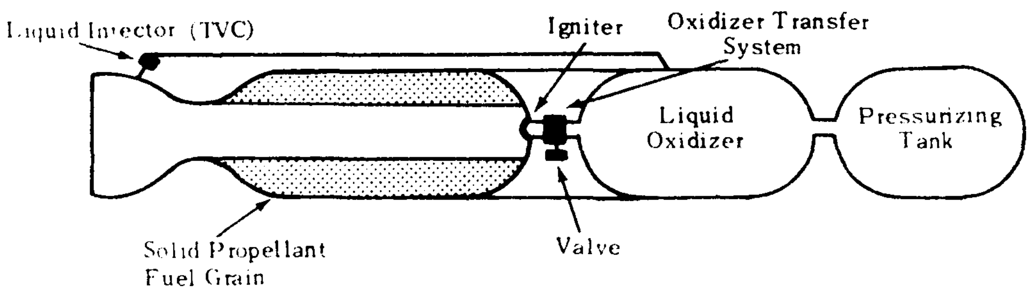


Figure 3-22. Hybrid rocket engine.

ADVANCED PROPULSION TECHNIQUES

As mission requirements approach or exceed the limits of chemical rockets, new propulsion techniques must be investigated. The new techniques will obey Newton's laws just as the contemporary techniques do, but they will use different energy sources and hardware to produce the propulsive force. This force will still be the reaction of the vehicle to mass being ejected at high velocity.

Need for Advanced Designs

The need for advanced designs becomes readily apparent as engineers consider velocity limitations of chemical rockets. Figure 3-23 shows these limitations for single and multistage vehicles. The graphs include both $LO_2/RP-1$, as well as the higher energy combination LO_2/LH_2. Payload fraction is the payload weight divided by the gross weight of the vehicle.

Since there is a practical limit to the number of stages that researchers can use to increase mass ratios, consider that I_{sp} is proportional to T_c/m. Remember also that T_c and m interact and influence each other in a combustion process. If T_c and m were independent, scientists could vary them so that the resulting ratio would be higher. The value for m is about 20 in present chemical engines.* If they reduce m, a higher performance engine results.

Nuclear Rocket

One program that has followed this approach is the nuclear rocket. It has increased the theoretical I_{sp} to more than 800 seconds—double that of current chemical engines. Figure 3-24 shows a schematic reproduction of a typical nuclear, solid core, thermal reactor engine. Since liquid hydrogen is the propellant, thrust levels are comparable to upper stage chemical engines using hydrogen.

Initially the nuclear rocket was a joint Atomic Energy Commission-USAF project but became an AEC-NASA program of two phases. The early reactor core design phase, using the Kiwi A and B reactors, reached completion in September 1964. The second phase involved developing the flight configuration engine known as the nuclear engine for rocket vehicle application (NERVA). The US Air Force planned preliminary flight tests for ballistic trajectory flights with the NERVA functioning as an upper stage main engine. Many areas of study and research are necessary before the nuclear rocket is ready for production. Among these are starting at altitude, and radiation and neutron heating that may affect the payloads, propellants, and structural materials. Engineers and Scientists have ground tested a more advanced increased density reactor called Phoebus, and they have considered a flyable engine configuration. However, the government has not funded this research and there is no work being done.

One theoretical improvement is a high-density reactor using fast neutrons. This type of reactor is expected to produce higher performance levels in a smaller package than the thermal (or slow) reactors mentioned above. Another improvement that may prove feasible at some time in the future is a gas core reactor, in which the operating temperature (T_c) could be much higher. This increase in temperature would occur because of the elimination of the solid core or fuel elements used in slow and fast reactors. These structural elements are temperature limited.

Low Thrust Rockets

Chemical engines produce high thrust for a short time (minutes), and nuclear reactor engines yield high thrust for hours. Each produces acceleration that stops when the propellant is exhausted.

There are other engines that produce only small amounts of thrust, but they do so for months, or even years. Thrust acting for long periods can produce final velocities much higher than those produced by chemical or reactor engines.

---

*The molecular weight for the *Space Shuttle main engine* (SSMF), which uses LO_2/LH_2, is about 10 to 12 at maximum I_{sp}.

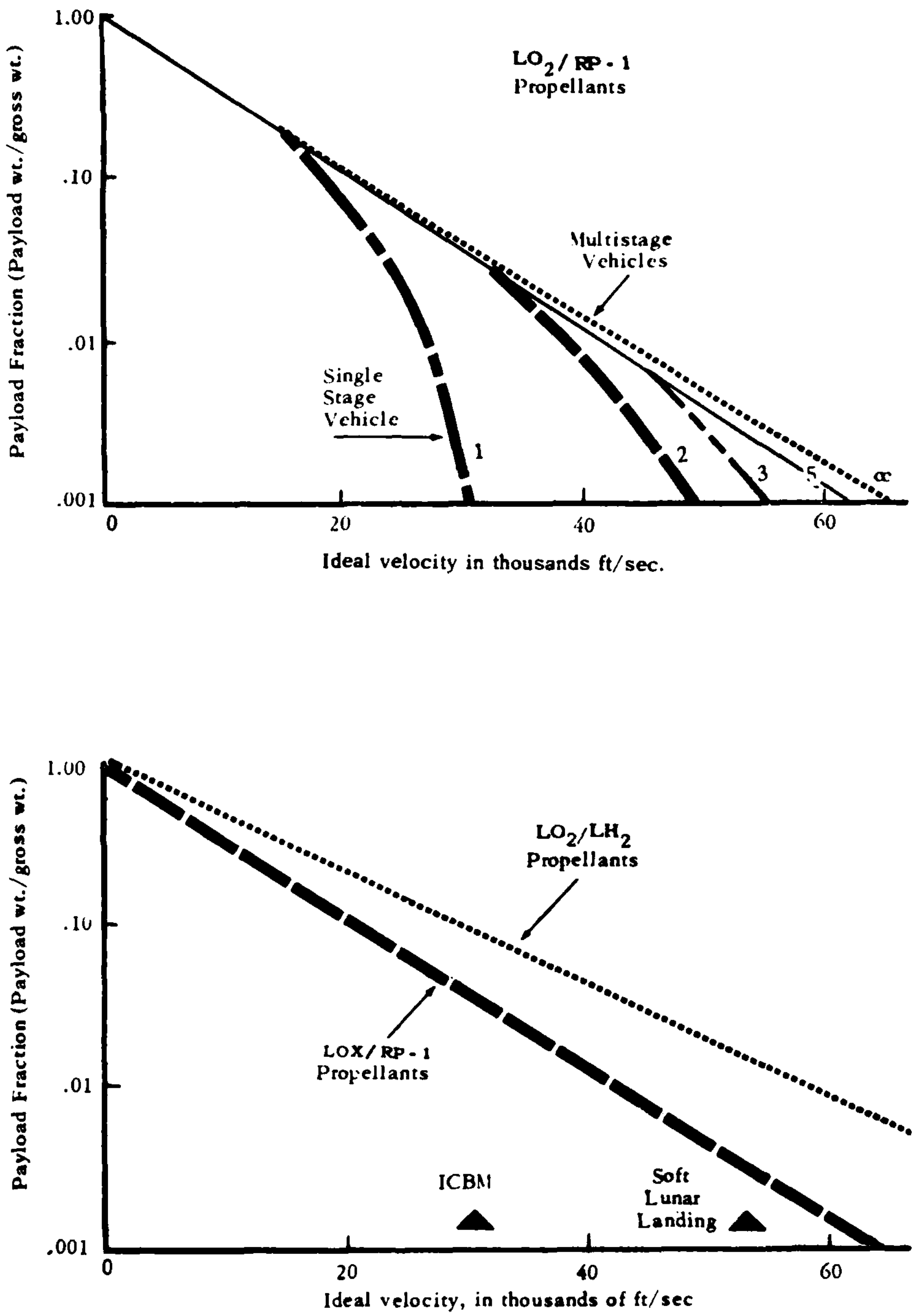


Figure 3-23. Payload fraction versus ideal velocity

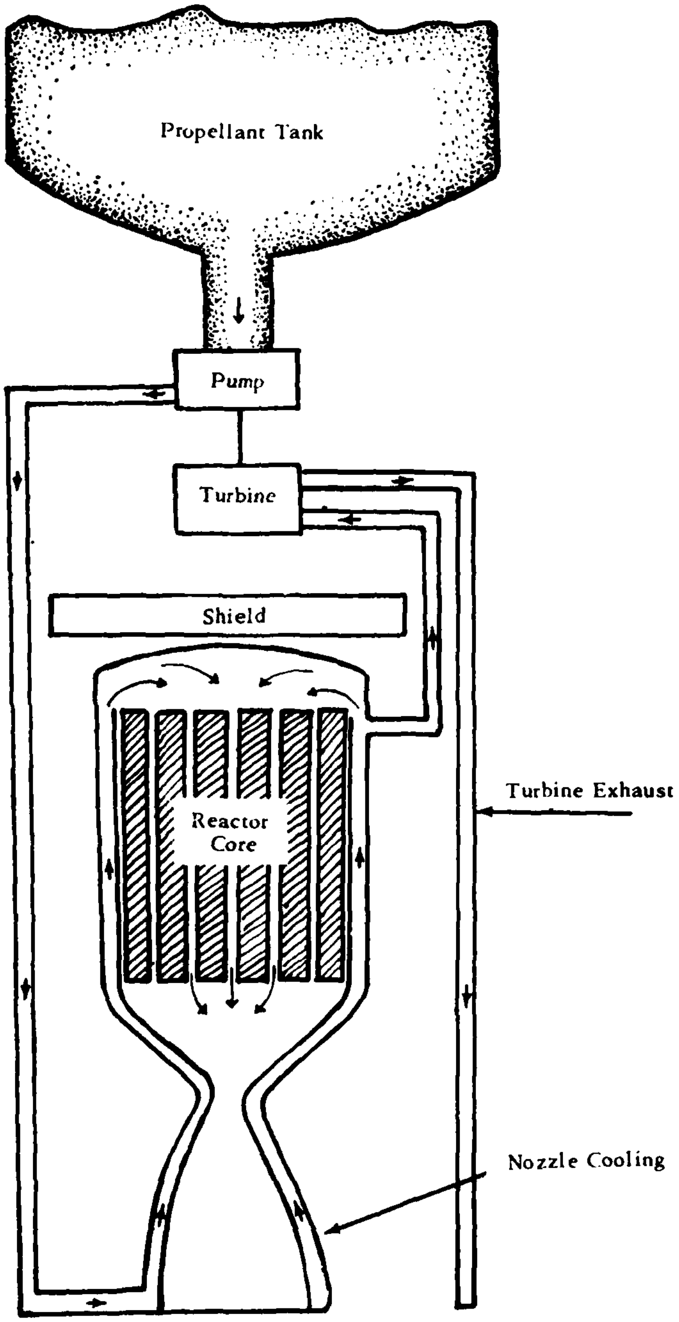


Figure 3-24. A nuclear rocket.

The *radioisotope heat cycle* and the *electric* engines produce thrusts measured in micro and millipounds, specific impulses from 700 to 30,000 seconds, and operating times ranging from days to years. These engines cannot lift themselves from the earth, but they can move large payloads through space.

The *radioisotope heat cycle engines* use high-energy particle sources such as plutonium and polonium. The walls of the isotope container stop the particles, thereby converting their kinetic energy to heat. The engine uses this heat to raise the temperature and pressure of a propellant that it expels through a nozzle.

The Decomposed Ammonia Radioisotope Thruster (DART) is such a device. Here 238 plutonium dioxide heats the ammonia to produce 0.1 pound of thrust. Scientists are seeking an operating time of at least one year.

Basically there are three types of electric engines: arc jet, ion engine, and plasma jet. The arc jet uses an electric arc to heat the propellant. The engine accelerates the propellant thermally and ejects it as a high-velocity plasma from a conventional nozzle. The ion engine is an electrostatic device that removes electrons from the propellant atoms to form positive ions. The engine electrostatically accelerates these ions and ejects them to produce thrust. Then, the engine must add electrons to make the exhaust electrically neutral to prevent accumulation of a negative charge on the vehicle. The plasma jet uses electromagnetic force to accelerate and eject the propellant in a plasma form to provide thrust. Electromagnetic force is necessary, since a plasma is a substance that is ionized but is electrically neutral (a plasma consists of an equal number of positive and negative ions, and therefore it has no net charge).

This is a very simplified presentation. The theory and associated equations for these devices, particularly those for the plasma jet, become quite complex.

The three basic types of electric engines have many subdivisions based on such considerations as design, the method of transferring energy to the propellant, thrust, propellant consumption rates, and I_{sp}. Scientists expect electric rockets to yield specific impulses of 2,000 to 30,000 seconds or more.

These engines have great potential for some missions. The concept receiving the most attention to date is the ion engine.

Ion Propulsion Theory

Figure 3-25 shows the principal elements of the ion-thruster propulsion system. The sun provides energy, which solar cells convert into electric power. Then the system conditions the power to the current and voltage that the ion thruster needs. It ionizes the propellant in the thruster and electrically exhausts the propellant to produce thrust. For many missions, the power source could serve the dual roles of providing both thruster power and power for mission objectives subsequent to the thrusting period. The thruster will be of appropriate size to satisfy the thrust requirements for the particular propulsion task.

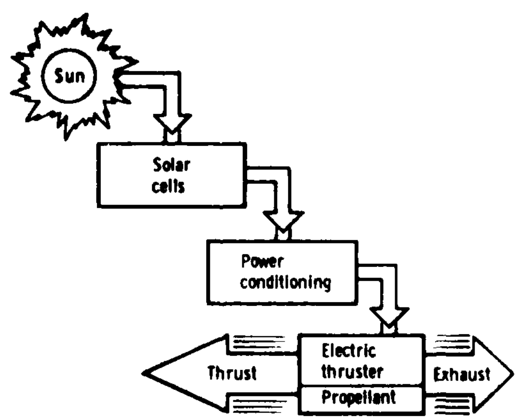


Figure 3-25. Ion thruster propulsion system.

The main advantage of using electric propulsion is that the electric energy added to the exhaust propellant greatly increases its velocity, or specific impulse, hence electric propulsion produces more thrust with the same propellant flow rate. Figure 3-26 shows the mass of propellant required to produce a given thrust decreases with increasing specific impulse. However, the increasingly massive power plant required to accelerate the exhaust to higher velocities offsets the saving in propellant mass. Figure 3-27 shows this increase in power plant mass. The spacecraft achieves the maximum payload at the optimum specific impulse, where the sum of the propellant and power plant mass is a minimum. Figure 3-28 shows this.

Figure 3-28 indicates that at low specific impulse the propellant mass can be excessively large, while at high specific impulse the power plant mass becomes excessive. Between these two extremes is a broad useful range where sufficient payload remains for design of a practical spacecraft. As defined by figure 3-28, payload includes the mass of the spacecraft itself and the useful payload. The optimum value of specific impulse to maximize payload usually is between 2,000 and 5,000 seconds, and thus the optimum value of exhaust velocity is between 20,000 and 50,000 meters per second. The spacecraft easily achieves this range of exhaust velocity with ion thrusters and, as we discuss later, this range results in strong increases in spacecraft payload over a large variety of missions.

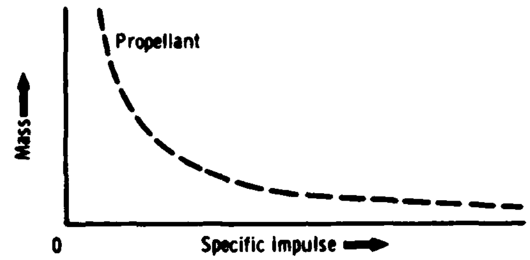


Figure 3-26. Effect of propellant mass on specific impulse.

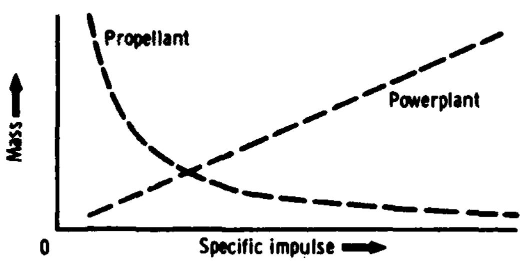


Figure 3-27. Mass/specific impulse with powerplant added.

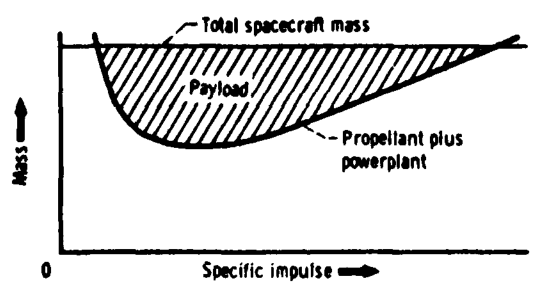


Figure 3-28. Mass/specific impulse with total spacecraft.

Ion-thruster operations. Dr. Harold R. Kaufman conceived and tested the first electron-bombardment thruster in 1959 at the NASA Lewis Research Center. This thruster operates by flowing a gaseous propellant into a discharge chamber. The propellant may be any gas, but mercury, cesium, and the noble gases are the most efficient for propulsion applications. Electronic bombardment ionizes propellant atoms in the discharge chamber in a process similar to that in a mercury arc sunlamp. This ionization occurs when an atom in the discharge loses an electron after bombardment by an energetic (40-eV) discharge electron. The electrons and the ions form a plasma in the ionization chamber. The electric field between the screen and the accelerator draws ions from the plasma. Then, these ions are accelerated out through many small holes in the screen and accelerator electrode to form an ion beam. Figure 3-29 shows this. A neutralizer injects an equal number of electrons into the ion beam. This beam of electrons allows the spacecraft to remain electrically neutral and is a requirement for successful thruster operation.

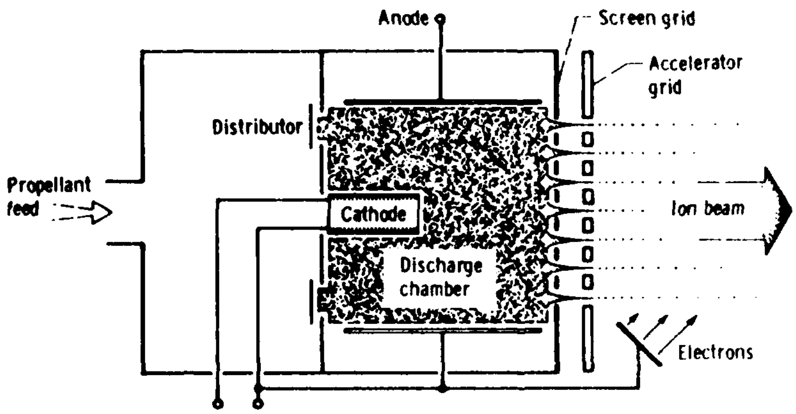


Figure 3-29. Ion thruster operation.

The development of the mercury-bombardment thruster has continued through the years since the 1960s. Scientists have successfully tested thrusters 2.5 to 150 centimeters in diameter. These thrusters require power of 50 watts to 200 kilowatts and produce thrust of 0.4×10^{-3} to 4 newtons (0.1×10^{-3} to 1 pound).

Many laboratories in the United States and Europe have worked on a wide variety of electric thrusters. These include colloid thrusters using a doped-glycerine propellant, a pulsed-plasma thruster using ablation of a Teflon propellant block, and a combardment thruster using cesium propellant. In England, France, and Germany numerous laboratories and universities are at work on electric thrusters for both auxiliary and primary propulsion. The electric propulsion effort by the Soviet Union includes flights of Zond, Meteor, and Yantar spacecraft with ion-thruster experiments onboard. In the United States, NASA has flown two spacecraft specifically to test ion thrusters in space. We discuss these tests, SERT I and SERT II, in the next two sections.

SERT I. Even though large vacuum tanks provide an excellent simulation of an environment for testing ion thrusters, scientists can answer some questions only by operating ion thrusters in space.

On 20 July 1964, NASA briefly tested two ion thrusters in space. One was a mercury electron-bombardment thruster developed by the NASA Lewis Research Center. The other was a cesium contact-ionization thruster developed by the Hughes Research Laboratories under NASA contract. The name given this test was Space Electric Rocket Test I (SERT I). Scientists mounted the battery-powered thrusters on a capsule that they launched with a Scout solid-propellant rocket into a ballistic trajectory.

The primary purpose of SERT I was to demonstrate neutralization in space and measure any differences between ground and space operation. The direct evidence of incorrect neutralization would be a decrease in thrust from the predicted values.

The cesium thruster onboard SERT I did not operate, but on August 29, 1964, a cesium thruster produced thrust in a similar space test conducted by the US Air Force. Electro-Optical Systems of Pasadena, California, developed the Air Force test thruster. Scientists successfully tested small electrothermal, resistojet thrusters in space in the late 1960s. They had tested small pulsed-plasma thrusters as well. Thus, they established that electric thrusters would work in space and that they could predict the thrusters' performance in space from ground-based tests in vacuum facilities. The next step was to determine how durable the thrusters were. As mentioned previously, electric thrusters produce only a small amount of thrust; therefore, they must be capable of operating for long periods—several months to years in most applications.

SERT II. The SERT I flight verified the neutralization of an ion beam in space by showing the production of thrust, but it was a short flight using batteries for power. The purpose of the SERT II flight was to demonstrate long-term operation of an ion thruster in space with a flight-type power source.

On 3 February 1970, a Thorad/Agena launch vehicle launched the SERT II spacecraft into a circular orbit (fig. 3-30). The polar orbit permitted the solar panels, which powered the thruster, to remain in continuous sunlight throughout the entire orbit.

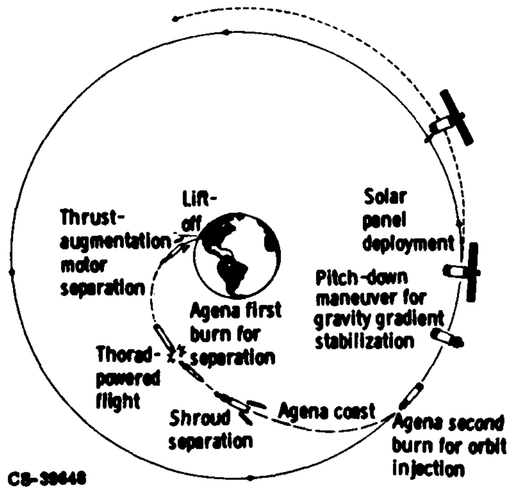


Figure 3-30. Flight of Sert II.

Each thruster (fig. 3-31) was a 15-centimeter-diameter mercury-bombardment ion thruster and at full power used 850 watts to produce 28 millinewtons of thrust. The thruster was able to operate at 40 and 80 percent of full power.

Results of the SERT II flight include five months of successful operation with one thruster and three months with the other thruster. Engineers believe a minor thruster redesign as a result of follow-on ground tests will provide future thrusters with an operating lifetime of 15 to 30 months. They stopped ground life tests of identical thruster-power processor systems after several months of continuous running. The tests were without failure. An onboard accelerometer and the change in spacecraft orbit measured the thrust of the SERT II ion thruster. These two measurements of thrust and a thrust calculated from measured beam current and voltage produced identical values of thrust within experimental error.

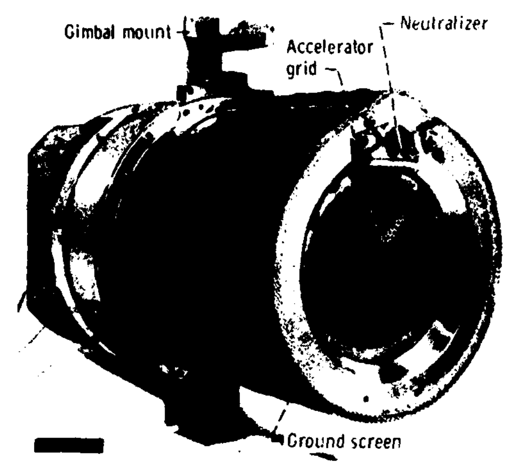


Figure 3-31. Sert II thrusters.

The SERT II spacecraft and thruster system components remained in good operation after several years in space. The spacecraft was able to start successfully two propellant flow systems and two hollow cathodes on each of the two flight thrusters over 200 times in space. Storage periods between starts ranged from 10 minutes to 350 days. Each thruster power processor operated without incident for 4,000 hours in space. The spacecraft solar arrays and thermal control surfaces showed no abnormal degradation due to contamination from thruster operation. The spacecraft attitude-control system and thruster gimbal actuators functioned correctly.

The SERT II flight has provided mission planners with important data needed to design space electric propulsion systems for long-duration missions. These data, combined with data from over 200,000 hours of ground tests of thrusters, provide a confident basis for evaluating thruster operating lifetime and thrust performance level and for designing future spacecraft to avoid contamination.

Auxiliary propulsion. Geosynchronous spacecraft often require careful control of their position and orientation over long periods of time. This is especially true of the highly directional antennas presently in use. Electric propulsion is ideal for the station keeping and attitude-control functions of geosynchronous satellites because of its specific impulse, which results in considerable propellant weight savings over chemical propulsion systems.

The most important application of auxiliary-propulsion ion thrusters is north-south station keeping. Figure 3-32 shows the tendency of the gravitational forces of the sun and moon to increase the inclination of the geosynchronous orbit. Figure 3-32 indicates that, through the use of proper thrusting centered about the nodal crossings, the geosynchronous orbit will not incline and will remain in the equatorial plane. The duration of thrust each day depends on the spacecraft mass and thrust level and on how closely engineers can align the thrusters with respect to a north-south line and ensure that the thrusters still have a thrust direction through the center of mass of the spacecraft. Figure 3-33 shows typical thruster orientations on a spinner spacecraft. Three-axis stable-platform spacecraft allow a direct north-south alignment of the thrusters if mounted as figure 3-34 shows.

Effects of solar pressure and the triaxiality (gravitational nonuniformity) of the earth will cause the east-west position of a geosynchronous spacecraft to vary unless the spacecraft applies thrust corrections properly. The triaxiality of the earth is a much weaker disturbing force than that caused by the sun and the moon and requires only 1/26 of the total impulse for proper correction. The thrusters (fig. 3-34) can easily perform east-west station keeping through the use of the thruster gimbal system. Their thrust is deflected in the correct direction and they are started simultaneously

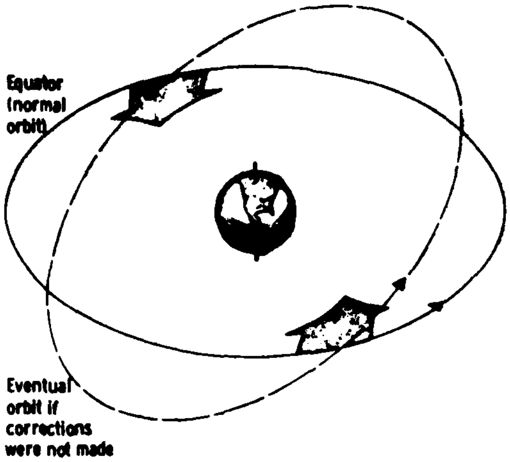


Figure 3-32. Effects of gravitational forces on geosynchronous orbit.

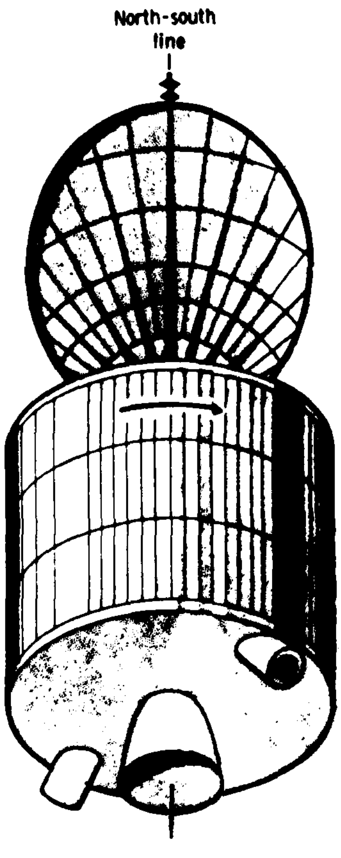


Figure 3-33. Typical thruster orientation on spinner spacecraft.

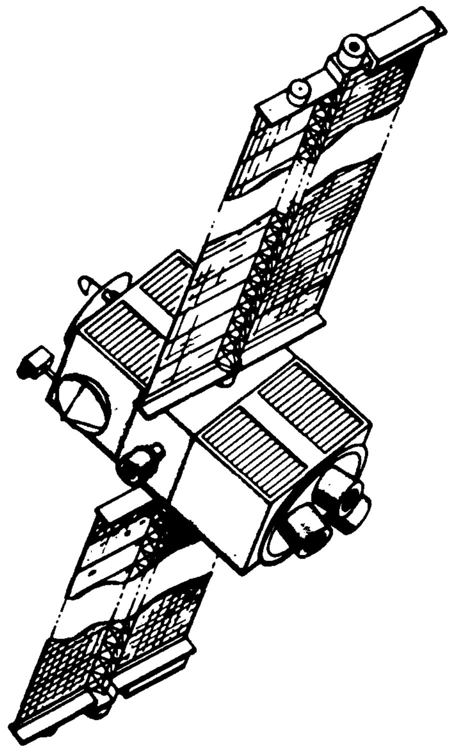


Figure 3-34. North-South thruster alignment on three-axis stable platform spacecraft.

at specific times throughout the orbit. Station walking, or changing the east-west location of the geosynchronous spacecraft can also be accomplished.

Because the center of solar pressure and the center of mass of the spacecraft are rarely the same point, the thrusters must increase gradually the speed of fly wheels to hold the spacecraft in proper orientation. Electric thrusters can provide a counteracting disturbing torque, which can be used to dump or reduce the velocity of the momentum wheels. This reduction can help lower excessive momentum-wheel spin rates. Figure 3-35 shows how the thrusters can help despin the momentum wheel for one of the three axes of a spacecraft.

Auxiliary electric propulsion can present a profit when used in place of other types of auxiliary propulsion for geosynchronous spacecraft. Figure 3-36 compares weights of a mercury ion-thruster system for north-south station keeping with those for a hydrazine system. As we can readily see, the use of electric propulsion can provide a significant weight saving on a long mission life spacecraft. A typical 347-kilogram (765 pounds) geosynchronous seven-year spinner spacecraft could save 51 kilograms (112 pounds) by substitution of mercury ion thrusters for a hydrazine north-south station keeping system.

Primary propulsion. Electric propulsion for primary spacecraft thrust is of interest for both near-earth and interplanetary missions. Near-earth applications include spiralout maneuvers from low to high orbit and earth escape (fig. 3-37). Once at high orbit, the spacecraft may use the thruster for station keeping applications. Interplanetary missions include flights out of the ecliptic and flyby past or rendezvous with asteroids, comets, and planets. A primary electric propulsion stage could offer large payload advantages as a commercial tug in conjunction with the space shuttle.

The interest in electric propulsion derives mainly from the reduction in propellant requirements

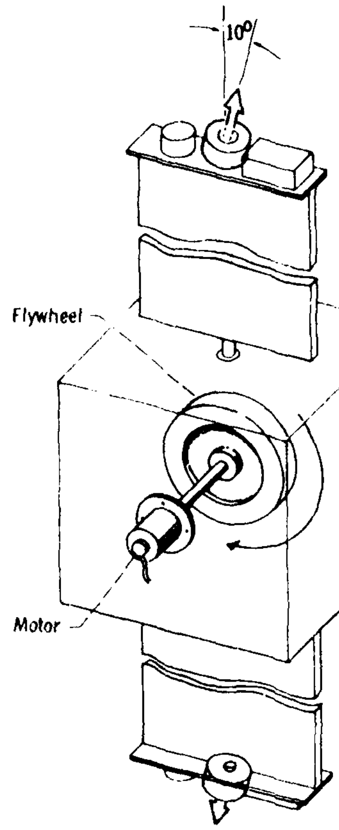


Figure 3-35. Auxiliary electric propulsion system for geosynchronous spacecraft.

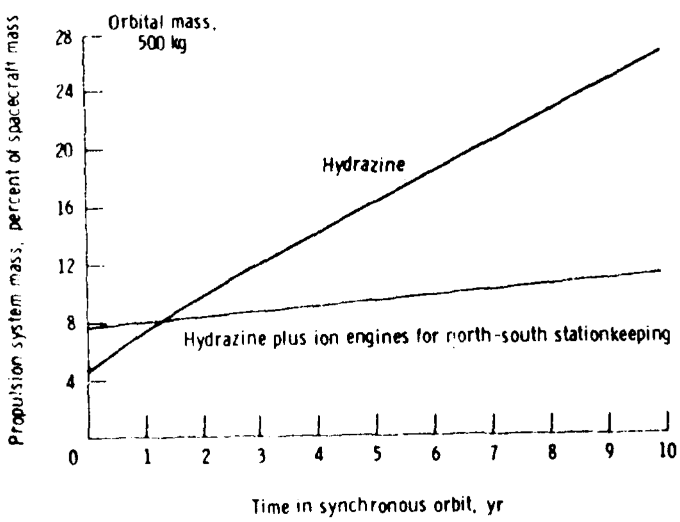


Figure 3-36. Mercury versus hydrazine weights.

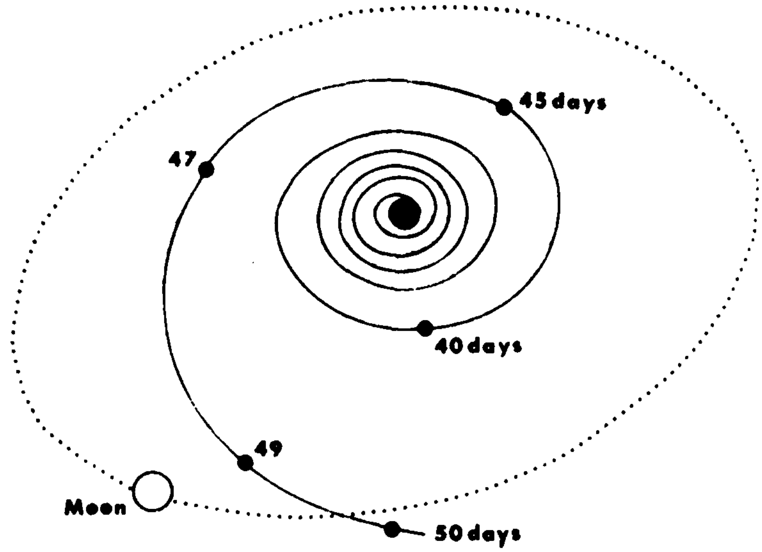


Figure 3-37. Escape trajectory.

relative to chemical propulsion due to operation at increased specific impulse. One way to compare the capability of an electric propulsion spacecraft with that of a chemical system is to consider the total impulse delivered by two such systems. A 1,500-kilogram electric propulsion spacecraft with 500 kilograms of propellant can deliver slightly more total impulse than 2914 Delta rocket stage, which has a mass of 5,500 kilograms, including 4,500 kilograms of propellant. Figure 3-34 shows a typical electric propulsion spacecraft. Although exact comparisons are subject to details of the propulsion system configuration, the comparison just given illustrates the propellant savings, and hence, overall mission performance increases, achievable with electric propulsion. A discussion of the major elements of a primary electric propulsion system follows to explain more fully its characteristics.

Intensive study and development have been under way at the NASA Lewis Research Center, the NASA Jet Propulsion Laboratory, Hughes Research Laboratory, and the NASA George C. Marshall Space Flight Center in the area primary propulsion applications. The candidate thruster for use on all proposed missions is the 30-centimeter thruster. This thruster operates at a nominal input power of 2.75 kilowatts at a thrust of 0.135 newton and a specific impulse of 3,000 seconds. The engine achieves thruster efficiencies in excess of 0.71 at full thrust, but there is some decrease in throttle conditions. The thruster has been designed to throttle over more than a 4:1 range of input power. It has a design goal lifetime in excess of 10,000 hours.

The thruster is qualified both mechanically and thermally and is compatible with the launch environment of boosters ranging from Thor/Delta to Titan with a variety of upper stages. It is able to operate in the thermal environment associated with in-bound and out-bound inter-planetary missions ranging from about 0.5 to 4 astronomical units. As verified in the SERT II flight, this thruster is capable of very long-term space storage and has highly reliable multiple-restart capability.

One arrangement generally useful on a primary propulsion subsystem is the use of common tankage for all thrusters. This arrangement requires a flow line device that can isolate electrically the thruster from the common propellant tankage system. Scientists have developed such an isolator, and many component and thruster life tests have indicated lifetimes well in excess of the 10,000-hour design goal.

Power processors for primary propulsion thrusters have been under development for several years. Early tests defined the power supply characteristics and control logic systems required to operate the thruster over large ranges of power and thrust. Processors capable of operation in a vacuum have

been developed. These processors have electrical efficiencies in the range of 0.85 to 0.91 over the thruster-operating envelope and scientists may interface them with a spacecraft computer. They may fabricate and qualify the power processor with an optimized configuration to provide all the functions required by primary propulsion missions.

Thrust vector control of the 30-centimeter thruster will be available to provide spacecraft attitude-control and station keeping functions. One- or two-axis gimbaling of the thruster individually or in combination with other thrusters will achieve this control. This concept allows the incorporation of the development of the thruster as a module into a specifically designed spacecraft altitude control system. Figure 3-38 shows a two-axis thruster gimbaling system. Scientists can use other concepts such as gimbaling sets of thrusters.

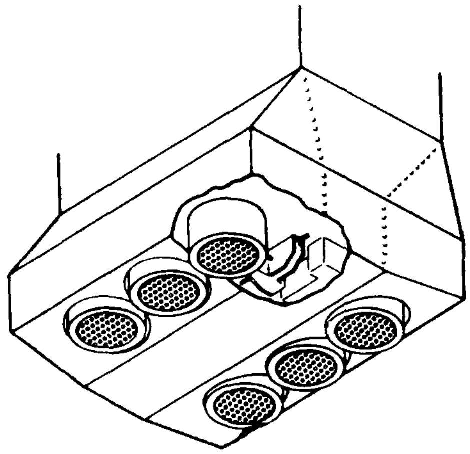


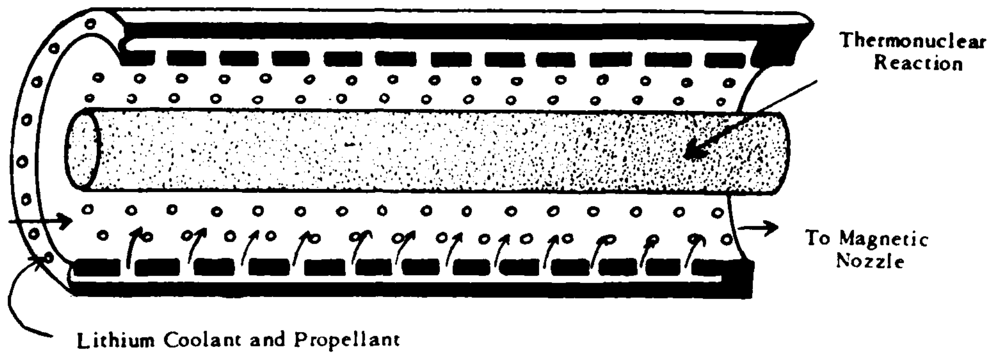
Figure 3-38. Two-axis gimbaling system.

The propellant supply and distribution system proposed for all missions is quite similar to that successfully flown on the SERT II mission. A common propellant supply, in one or more tanks, provides the propellant to all thrusters. The engine turns each thruster on or off as required during the mission. A blowdown system provides the required pressure. Use of appropriate internal liners allows multiple-mission use for a single qualified tank design.

Fusion propulsion. One advanced concept would use fusion reaction contained by an electromagnetic field (fig. 3-39). The field heats and controls the reaction, and controls the ejection of mass from the exhaust nozzle. The propellants are two isotopes of hydrogen. Deuterium (H^2) and Tritium (H^3) are proposed because they can sustain a fusion reaction when sufficiently heated. This concept may yield a theoretical specific impulse of one million seconds.

Scientists have proposed using a porous envelope through which the engine would force lithium coolant (and propellant) into the reaction. The lithium would enter into the reaction and the engine would expel it with the other exhaust products. When the engine does this, more mass flows, lowering the specific mpulse to a value of about 4,000 seconds, but the thrust level rises to a very high value.

Laser propulsion. NASA is investigating the use of laser beams for spacecraft propulsion or for auxiliary power. Several concepts are under study, but the basic scheme involves transmitting the laser energy from some platform (ground, airborne, or orbiting) through an optical window to heat a working fluid. This fluid could be hydrogen, perhaps seeded with cesium or carbon to improve energy absorption. Very high temperatues and high I_{sp} would be possible, over 1,000 seconds. Potential applications include perigee propulsion, so called because the spacecraft would be thrusting near



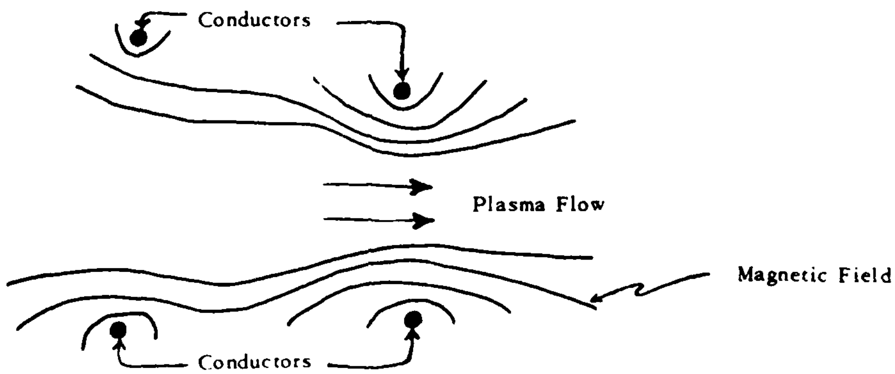


Figure 3-39. Fusion propulsion concept.

perigee to gradually raise the apogee over several orbits. Other applications might be orbit circularization, station keeping, or injection into a planetary transfer trajectory. Work is underway to study application concepts, transmission technology, and development of various lasers (particularly CO_2 lasers), as well as the technical areas of energy transfer and window materials.

SUMMARY

This chapter presented the basic laws of rocket operation, and some of the terminology and definitions pertaining to rocket engine performance. It discussed the interaction of design and sizing parameters of rocket vehicles, and methods of increasing rocket vehicle performance. We reviewed some of the advance propulsion concepts as well.

A comparison of the various propulsion devices on the basis of specific impulse alone does not present the complete picture. Only chemical propellants and high-thrust nuclear rockets have thrust-to-weight ratios that are sufficient to permit launches from the earth or from other bodies, such as the Moon, Mars, and Venus, that have gravitational fields. Others (electric) are applicable to missions in low-gravitational fields. These fields could be in deep space or between one planet's orbit and another planet's orbit. Again we would need a vehicle with chemical or nuclear reactor engines to travel from the orbital vehicle to a planet or moon surface and to return to the orbital vehicle.

PROPULSION SYMBOLS

a—acceleration
A_e—nozzle exit area
A_t—nozzle throat area
C_p—constant pressure specific heat
C_v—constant volume specific heat
F—thrust
g—acceleration of gravity
I_d—density impulse
I_{sp}—specific impulse
I_t—total impulse
k—ratio of specific heats
$\ln$—natural logarithm
m—molecular weight
M—mass
MR—mass ratio
p—electric power
P_c—combustion chamber pressure
P_e—exhaust pressure
P_o—ambient pressure
Q—reactor thermal power
r—mixture ratio
t_b—burning time
T_c—combustion chamber temperature
T_e—exit temperature
T_i—inlet temperature
V—volume
Δv_a—actual velocity change
Δv_i—ideal velocity change
Δv_l—velocity change losses
v_e—nozzle exit velocity
v_r—earth surface velocity
W—weight
$\dot{W}$—weight rate of flow
$\dot{W}_f$—weight rate of fuel flow
$\dot{W}_o$—weight rate of oxidizer flow
W_p—propellant weight
W_1—vehicle weight at engine start
W_2—vehicle weight at engine shutdown
ϵ—nozzle expansion ratio
Ψ—thrust-to-weight ratio
η—electrical efficiency
ν—heat transfer efficiency

SOME USEFUL PROPULSION EQUATIONS

1. Thrust of rocket: $F = \frac{\dot{W}}{g} v_e + A_e (P_e - P_o)$
2. Expansion ratio: $\epsilon = \frac{A_e}{A_t}$
3. Measured specific impulse: $I_{sp} = \frac{F}{\dot{W}_p}$
4. Mass ratio: $MR = \frac{W_1}{W_2} = \frac{\text{Engine start weight}}{\text{Engine stop weight}}$
5. Overall mass ratio: $MR = (MR_1) \times (MR_2) \times (MR_3) \times \ldots$
6. Thrust-to-weight ratio: $\Psi = \frac{F}{W}$
7. Lift-off acceleration: $a = (\Psi - 1)$ g's
8. Ideal velocity change: $\Delta v_i = I_{sp}\, g \ln MR$
9. Actual launch vehicle velocity: $\Delta v_a = \Delta v_i - \Delta_l + v_r$
10. Theoretical specific impulse:

$$I_{sp} = 9.797\sqrt{\left(\frac{k}{k-1}\right)\left(\frac{T_c}{m}\right)\left[1 - \left(\frac{P_e}{P_c}\right)^{\frac{k-1}{k}}\right]}$$

11. Mixture ratio: $r = \frac{\dot{W}_o}{\dot{W}_f}$

12. Density impulse: $I_d = (I_{sp}) \times (SG)$

13. Total impulse: $I_t = (F) \times (t) = (I_{sp})\ (W_p)$

14. Total Δv (multistage vehicle): $\Delta v_t = \Delta v_1 + \Delta v_2 + \Delta v_3 + \ldots$

BIBLIOGRAPHY

Billman, Kenneth W. "Laser Energy Conversion." *Astronautics and Aeronautics* (July-August 1975):56.

Clark, K. E. "Survey of Electric Propulsion Capability." *J. Spacecraft* 12 (November 1975).

Dipprey, Duane F. "Liquid Propellant Rockets." *Chemistry in Space Research.* Robert F. Landel and Alan Rembaun, ed. New York: American Elsevier Publishing Co., 1972.

Finke, R. C. *Current NASA Ion Propulsion Technology.* NASA Report TMS-71612, October, 1974.

Hesse, W. J. and N. V. S. Mumford, Jr. *Jet Propulsion for Aerospace Applications.* New York: Pitman Co., 1964.

Lewis Research Center. *Electric Propulsion for Spacecraft.* Cleveland, Ohio, Lewis Research Center (NASA), 1974.

Mead, F. B., ed. *Advanced Propulsion Concepts-Project Outgrowth.* Edwards AFB, Calif., Air Force Rocket Propulsion Laboratory, AFRPL-TR-72-31, June 1972.

Minovitch, M. A. *Performance Analysis of a Laser-Propelled Interorbital Transfer Vehicle.* NASA Report CR-134966, February 1976.

Papilou, D. D. *Frontiers in Propulsion Research: Laser, Matter-Antimatter, Excited Helium, Energy Exchange, Thermonuclear Fusion.* NASA Jet Propulsion Laboratory, TM 33-722, 15 March 1975.

Pedersen, E. S. *Nuclear Energy in Space.* Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Schwenk, F. C. "Propulsion for Space: The Nuclear Rocket." *TRW Space Log* 7 (Spring 1967):2-24.

Shapiro, A. H. *The Dynamics and Thermodynamics of Compressible Fluid Flow.* New York: The Ronald Press Co., 1953.

Stuhlinger, E. *Ion Propulsion for Space Flight.* New York: McGraw-Hill, 1964.

Sutton, G. P. *Rocket Propulsion Elements.* New York: Wiley and Sons, 1964.

Yafee, Michael L. "Lasers Investigated for Space Propulsion." *Aviation Week and Space Technology* (21 April 1975):47.

Zucrow, M. J. *Aircraft and Missile Propulsion.* New York: Wiley and Sons, 1958.

Chapter 4

SPACE VEHICLE ELECTRICAL POWER

The availability of electrical power has been partially responsible for the successful operation of flight vehicles in space. In general, a space vehicle has neither utility nor purpose without some form of electrical power, which is the lifeblood of a space vehicle, vital to the needs of both passengers and equipment. These needs include power requirements for communications, data gathering and handling, vehicle attitude control, guidance and control, life support equipment, and many other areas. With few exceptions, all space vehicles without electrical power become useless debris. The space vehicle electrical power subsystem is, therefore, one of the most important spacecraft elements. Its design often dictates the active functions that a space vehicle can perform. Furthermore, in the future some space vehicles will use electrical power to energize electrical space propulsion systems.

Three government agencies that have researched space electrical power systems are the Air Force Geophysics Laboratory, the Air Force Aero Propulsion Laboratory, and NASA Marshall Space Flight Center.

In this chapter, we discuss three areas pertaining to space vehicle electrical power: electrical power generation in space, energy sources and associated energy conversion devices, and the evaluation and comparison of specific power subsystems.

PRODUCING POWER IN THE SPACE ENVIRONMENT

The space environment consists of a near vacuum (10^{-12} Torr at 7,000 nautical miles) containing both charged and uncharged particles; electric, magnetic, and gravitational fields; and electromagnetic radiation from both galactic and extragalactic sources. These and other environmental factors find the space power design and systems engineer working with vastly tighter constraints than the conventional power engineer. A discussion of some of the problems of producing electrical power includes the following areas:

1. Removal of waste heat. A critical and ever-present problem in space vehicles is the disposal of waste heat produced by power generation and power conditioning equipment. Earth-based systems for generating electrical power can use convection, conduction, and radiation to remove excess heat. However, the virtual absence of matter around a space vehicle makes radiation the only economical means of waste heat disposal. But, space is an inefficient heat sink; hence, large radiators must sometimes be employed, imposing severe weight and volume penalties. Besides these penalties, a large radiator increases the possibility of meteoroid damage.

2. Meteoroid bombardment. A possible vexing problem is potential catastrophic equipment failure resulting from puncture by meteoroids moving at relative velocities as high as 150,000 miles per hour. Although the probability of this occurrence is statistically quite negligible, we must consider the potential damage or erosion effects of meteoroid bombardment.

3. Very high vacuum. Operating space systems in a very high vacuum (which some require)

exposes them to problems of cold welding, outgassing, and leakage. Furthermore, rotating equipment requires lubricants that will not sublimate. Some components require a pressurized environment.

4. Radiation. Among the radiation hazards are cosmic rays, solar proton emissions, and intense ultraviolet and infrared radiation. However, the most serious radiation hazards of space flight to both equipment and people are the geomagnetic-trapped Van Allen radiation fields and the solar corpuscular radiation associated with solar flares. This radiation can cause serious damage to solar cells and semiconductors as well as people. Presently, shielding is considered the most suitable technique for protecting power systems from these radiation hazards.

5. Weightlessness. In earth-based electrical power generation systems, we use the force of gravity on a body (weight) in the separation of vapors from liquids in boiling and condensing fluids, in convective heat transfer, and in separation of gases from liquids. Since we use these various processes often in generating electrical power, systems cannot use components or processes that operate only with "this end up."

6. Temperature variation. Earth-based systems operate with relatively small temperature variations. In space the temperature of components within the power system may vary through extreme ranges because of solar heating and heat generated by the system itself. The power output of chemical batteries is very sensitive to temperature change. In general, the efficiency of a chemical battery decreases as its temperature decreases, but the output of solar cells increases as their temperature decreases. Also, the useful life of other electrical components depends on the operating temperature.

7. Launch environment. Earth-based systems for generating electrical power do not operate under adverse conditions. Engineers must design the delicate components of space systems to withstand the high acceleration and vibration of the launch phase.

8. Kinetic perturbations. Scientists have fixed the structure of earth-based systems for generating electrical power solidly to the earth to counteract any unbalanced forces developed by moving parts. Space vehicles will change attitude due to forces such as those produced by armatures rotating at high speeds or the movement of sun-oriented solar arrays. Counteracting these forces adds to the complexity of the flight vehicle.

Many other problems, combined with the areas we have already discussed, make the task of electrical power generation in space truly a very difficult one from an engineering design point of view. Among them are the limited or nonexistent repair capabilities; the very high premium placed on minimum weight, long life, and high reliability; and the inherent low efficiency of energy devices.

TYPES OF SPACE ELECTRICAL POWER SYSTEMS

By their energy sources, we can categorize the electrical power systems for space applications into areas of electrochemical, solar, or nuclear. We may categorize them further into static or dynamic energy conversion systems. Static systems of current interest include batteries, fuel cells, and photovoltaic, thermoelectric and thermionic devices. Dynamic energy conversion systems consist of either turbine or reciprocating engines coupled to electric generators. The engines would operate on thermal energy derived from either the sun, a nuclear reactor or isotopes, or the combustion of chemical fuels. Now we will discuss each of these systems in greater detail.

Electrochemical Systems

Batteries, fuel cells, and chemical-dynamic systems represent the electrochemical power systems. Of these, the batteries have played the most important role in furnishing space power until recently.

Virtually every space vehicle launched has contained one or more batteries. The reason for their use is that they are very reliable. Furthermore, they are cheap, simple, and available.

Batteries. There exist two types of batteries: primary cells and secondary batteries. Generally, we consider primary batteries as one-cycle batteries since they get no recharging after discharging. This type of battery converts chemical energy directly to electrical energy. The greatest utility of the primary battery as a source of electrical power is either in launch vehicles or in space vehicles that have very short missions. This type of battery is also useful for special applications such as providing pyrotechnic power to energize explosive bolts to separate vehicle stages.

The primary battery used most frequently in space applications and having the highest energy density of batteries in common use is the silver-zinc battery. It consists of silver-silver oxide and zinc electrodes immersed in a potassium hydroxide electrolyte. Its useful energy density is in the range of 20 to 100 watt-hours per pound depending upon the discharge rate. This means we must provide 10 to 50 pounds of batteries for each kilowatt-hour required. Research is being done on advanced primary cells using more energetic anode materials, such as lithium, magnesium, and aluminum, that may provide energy densities in the neighborhood of 150 watt-hours per pound.

However, primary batteries have a limited life, determined by the active composition materials. In general, primary batteries are used extensively for space electrical power requirements of a few milliwatts to approximately one kilowatt for periods less than a week. For example, the Mecury spacecraft of 1962 and 1963 received power from 150 pounds of silver-zinc primary batteries. Because the time in orbit of these craft was not very long, the use of primary batteries was adequate. Similarly, the lunar module received power from silver-zinc batteries as did the Apollo command module during the critical reentry and post landing periods.

Space vehicles that use solar-powered energy conversion systems need secondary (or storage) batteries if they require power during solar dark periods. Except for the initial charge, the chemical reaction within secondary batteries provides a means to store the electricity produced by the sun's energy rather than acting as a primary energy source. The most widely-used secondary battery in space applications is the nickle-cadmium battery. The advantages of this type battery in comparison to other secondary cells are its demonstrated superior reliability over large numbers of duty cycles, small voltage excursions, high-charge rate acceptances, low-temperature operation, and long shelf life. However, one disadvantage of the nickel-cadmium battery is its low-energy density. Typical energy densities are from 2 to 15 watt-hours per pound, depending upon the required lifetime and charging rate. Thus, for applications that do not require long lifetimes, we use either silver-cadmium or silver-zinc batteries quite often because of their higher energy densities.

New secondary batteries, which are actually hybrid battery-fuel cell devices, are now under development. These hybrids use conventional battery positive electrodes in combination with a fuel cell negative electrode. The result is higher weight energy density and potentially longer life. The volume energy density however is lower because of the increased cell volume required to store hydrogen gas at pressures of about 500 pounds per square inch of area. Nickel-hydrogen and silver-hydrogen batteries are examples.

Conventional cells using aqueous electrolytes are close to the limits of development. Consequently, scientists are directing their high-performance battery research to nonaqueous electrolytes, which permit the use of highly active electrode materials such as the alkali metals.

Current research is primarily with some fundamental problems of ambient temperature, non-aqueous electrolytes, and the problem of containing concerned chemical reactions with standard free energies in the 100 Kcal/mole range. The problem of direct chemical reaction of the constituents of the cell relates to the thermodynamics of the system and the rates of various reactions involved. The presence of trace impurities, especially water, is an important determinant of corrosion rates. Therefore, scientists have devoted much effort to purification and analysis techniques and the manipulation of solvents and solutions in a dry atmosphere. A major difficulty raised by almost all nonaqueous electrolytes is low conductance, which limits the discharge rate of a battery. Conductance mechanisms are immensely complex, involving such related properties as viscosity, dielectric

constant, molecular association, and ion-solvent interactions. Scientists have measured the dielectric properties of a large number of electrolytes in various organic solvents including propylene carbonate, butyrolactone, and dimethyl sulfoxide. The effects of electrolyte type and concentration and the frequency dispersion of the dielectric constant provide insight into the structure of the solution. Other problems include the mechanism of the lithium electrode reaction under various conditions and finding viable positive electrodes. The lithium electrode is generally satisfactory in a clean system. Transition metal halides are generally unsatisfactory because of their solubility and high sensitivity. However, the recent development of inorganic electrolyte cells wherein the positive electrode reactant is the solvent itself has circumvented this problem to a large degree.

Fuel cells. In recent years, scientists have done considerable research in developing space qualified electrochemical fuel cells. Unlike batteries, fuel cells use chemical fuels and oxidants stored externally to the cell. Figure 4-1 illustrates the principal parts of one common type of fuel cell. The cell has two sintered porous nickle electrodes immersed in a solution of sodium hydroxide or potassium hydroxide. There is an external supply of hydrogen and oxygen gases. The hydrogen and oxygen, at a pressure of about 50 atmospheres, diffuse through the porous electrodes of the fuel cell. Chemical reaction between the hydrogen and the solution, and the oxygen and the solution, takes place on the electrodes. Ions migrate through the solution, and electrons flow through the external circuit.

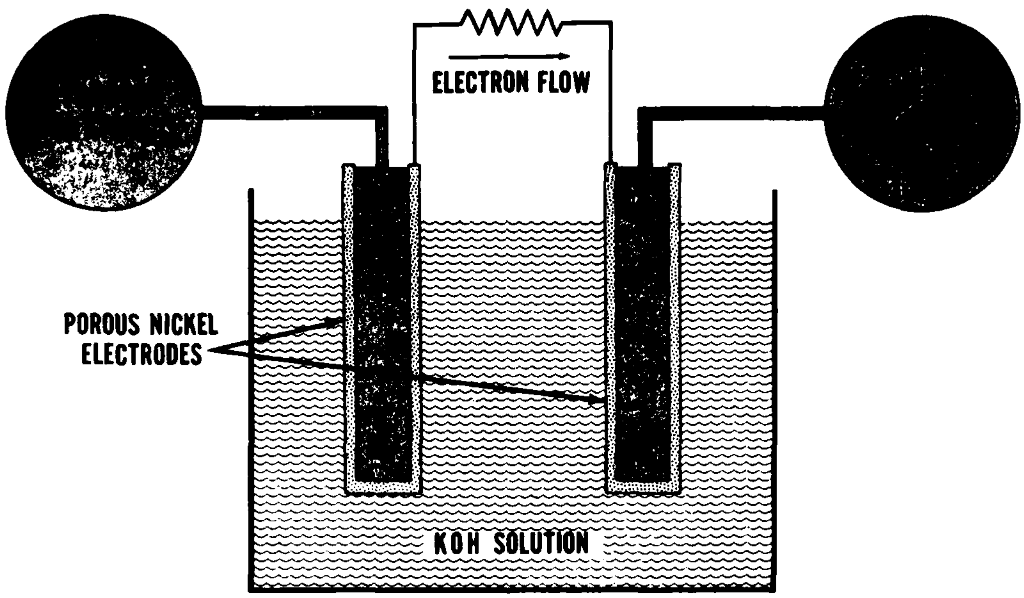


Figure 4-1. Schematic drawing of a hydrogen-oxygen fuel cell.

Scientists must ensure that the cell receives the chemical reactants constantly while operating and they must ensure the removal of the reaction product. The reaction product of the hydrogen-oxygen fuel cell is potable water. Both the reactant consumption and the water produced is approximately 1 pound per kilowatt hour. Scientists consider the hydrogen-oxygen fuel cell particularly suitable for short-duration manned missions due to its high efficiency and production of potable water. The Apollo program used hydrogen-oxygen fuel cells as a primary source of space power and drinking water.

Generally, energy densities or specific weight figures for fuel cells are somewhat meaningless because their magnitude varies considerably according to the operating time. This is because the weight of any given fuel cell is practically invariant to the time of operation. On the other hand, the chemical reactant weight is directly proportional to the required operating time and electrical load. From a weight standpoint, fuel cells are highly competitive for applications requiring approximately 10 kilowatts or less of power for operating periods of a few days to three months. Power densities of 200 to 900 watt-hours per pound are within current fuel cell technology for such operational times.

Fuel cells can operate either as an open cycle system or as a regenerative, closed cycle, system. The open cycle system can yield high current but only for the time that the supply of hydrogen and oxygen lasts. Increased operating time necessitates increased supply reactants with an accompanying increase in the size and weight of the tanks. In a regenerative system, the system could use energy from some other source, such as solar cells, to produce hydrogen and oxygen from the water formed in the fuel cells. Such a system, if perfected, could have an operational lifetime of several years. However, only the open-cycle system has been qualified so far and regenerative fuel cells are not currently being considered for near term applications.

Chemical dynamic. Chemical dynamic systems convert the energy available from a chemically-fueled combustion engine into electrical power by mechanical means. Such a system falls at the lower end of the power level-duration selection curve because of its relatively high rate of fuel consumption, which is two to three times that of fuel cells. However, this type of system generates AC power at useful voltage and frequency, simplifying power conditioning requirements. For high-power systems, approximately 10 kilowatts and higher, it tends to be lighter, more rugged, and considerably less temperature sensitive than electrochemical systems.

Although chemically fueled combustion engines have been extensively exploited for ground and aeronautical applications, ballistic or space efforts have been limited primarily to systems development. The propellants most frequently considered are hydrogen-oxygen, hydrazine, and the broader category of solid propellants.

It had been planned to use chemically-fueled combustion engines in both the Navaho and Dinosoar programs. Although the government has terminated both of these programs, many space power engineers still believe that the chemical engine will have a weight advantage over the fuel cell for operational time requirements of less than two weeks.

Solar Powered Systems

One result of the fusion inferno taking place in the sun is the release of an enormous amount of energy. A considerable amount of this energy occupies the eletromagnetic radiation spectrum for approximately ¼ micron to 3 microns (or millionths of a meter). A small fraction of this energy lies in the invisible untraviolet region. About half of it is the visible light; the rest is infrared, which accounts for the sun's heat.

At a distance of one astronomical unit (AU), the average amount of solar energy, the "solar constant," is approximately 130 watts/ft² of surface aligned at right angles to the sun's radiation. At other than one AU, the energy available equals the inverse square of the distance in AU (between the sun and the point of interest) times the solar constant at one AU. Thus, in the vicinity of Saturn, almost 10 AU from the sun, the available solar energy is only approximately 1.30 watts/ft², which is 1/100 of the available energy just above the earth's surface.

There are basically two primary conversion concepts that have been considered seriously as a means of converting solar energy to space electrical power, photovoltaic and thermal. Of these, the scientists have given the principal attention to the photovoltaic concept. Presently this concept provides the basis for the most common source of electrical power for unmanned spacecraft.

Photovoltaic. The photovoltaic effect is the generation of a voltage by photons (the smallest quanta of light) incident on a properly-treated surface that lies near the junction of two layers

of somewhat different material. Several types of photovoltaic materials can be used to provide useful space electrical power. However, only doped single-crystal silicon solar cells have been used operationally for efficiency reasons. Typically, contemporary solar cells are thin wafers, either 1 × 2 or 2 × 2 centimeters in area by about 8 to 14 mils thick (without filters). Electrically, the cell is n on p construction (top contact negative) with a base resistivity of 10 ohm-centimeters. The contacts are evaporated silver and titanium. The voltage potential produced by each cell is approximately 0.5 volt. The cells, connected in series, supply the desired voltage. Then, these subunits, connected in parallel, supply the desired current to be drawn from the complete unit. The engineers can mount the solar cell subsystem either on panels, paddles, or directly on the skin (body-mounted) of the space vehicle.

One of the inherent weaknesses of the solar cell has been its susceptibility to lose effectiveness due to the nuclear and other radiation in space. To decrease this damage, a layer of quartz or sapphire covers each solar cell. This layer absorbs some of the harmful radiation. Also, each cell receives a reflective coating to cause the reflection of light wavelengths below approximately 4,000 angstrom units. The energy in this region has a small effect on the cell, and by reflecting the energy, the cells can operate at a slightly lower temperature, thus giving more power. Also an anti-reflective (SiO_2) coating is applied to the cell to reduce the loss from reflection at the desired wavelengths. In addition to increasing cell efficiency this anti-reflective coating provides the cell its characteristic blue color. The filter, reflective and anti-reflective coatings, and adhesives add another 4 to 8 mil thickness, yielding a typical total solar cell thickness of 12 to 22 mils.

Unmanned space missions have used solar power subsystems more extensively than any other power subsystem. One reason is that solar units do not have to carry any fuel as a source of energy, since the energy source is the sun. However, the production of any appreciable amount of power requires large surface areas. The present level of solar cell technology is 11 to 12 percent efficiency (at air mass zero) and a specific weight approximately 10 watts per pound. Thus, for example, after accounting for all losses and degradations, a fully-oriented silicon solar cell array can provide a specific power of approximately 10 watts/ft^2. Table 4-1 shows a list of solar array parameters for a selected number of spacecraft using typical solar cell mountings. Note that the body-mounted and fixed paddle solar cell subsystems have specific weights and powers considerably less than those of oriented solar arrays. This is because the space vehicle generates the maximum electrical power at any location only when it aligns its solar cells to face the sun directly. The solar power available at 1 AU when the vehicle does not align the solar cells is $P = 130 \cos i$ watts/ft^2, where "i" is the angle of incidence (the angle between the direction of the sun's radiation and the perpendicular to the solar cell surface). Furthermore, in the absence of the sunlight, there is no production of electricity.

This latter aspect is an inherent weakness in the use of solar cells in that storage batteries must supplement a solar array subsystem when it is used on earth-orbiting spacecraft that require continuous power. This is because such spacecraft generally pass into the earth's shadow at regular intervals. The percent of time spent in the earth's shadow can be as high as 40 percent for low-altitude orbits. The spacecraft requires storage batteries to provide power during the dark periods. The sun recharges the storage batteries after the spacecraft emerges from the darkness.

Although the fully-oriented solar panels have size and weight advantages, the vehicle does not obtain these without associated disadvantages. Primary among the latter are the difficulties that the fully-oriented solar arrays incur in their integration with the spacecraft. Since the craft often uses its attitude control subsystem to aim the antennas, it requires separate attitude positioning and a sun sensor to orient the solar arrays. Furthermore, to allow positioning and power transfer, the movable solar arrays require such components as bearings, slip rings, and rotary transformers. Another integration problem is to eliminate or minimize the shadowing of portions of the solar array by other spacecraft-projecting components (sensors and antennas) and vice versa. The deployment and orienting equipment also cause lower reliability.

In addition, for low-earth orbits, atmospheric drag and other perturbations exact a spacecraft stabilization fuel penalty for using any solar system other than a body-mounted subsystem. The sum of all of these penalties is such that generally the use of an oriented system at altitudes less than

Table 4-1
Selected Typical Spacecraft Solar Array Parameters

Spacecraft	*Launch Year*	*Total PWR (Watts)*	*Type Subsystem*[a]	*1x2 cm Cells/ Spacecraft*	*Watts/*[b] *Pound*	*Watts/ Sq Ft*
Vela	1963	90	Body Mounted	13,236	2.67	**2.26**
OGO A	1964	710	Oriented	32,256	6.79	**8.94**
Mariner-Mars	1965	680	Oriented	28,224	9.6	**9.7**
IDCSP	1966	46	Body Mounted	7,808	1.96	**2.04**
OGO D	1967	745	Oriented	32,256	7.12	**9.38**
Intelsat III	1968	167	Body Mounted	10,720[c]	5.20	**3.56**
Nimbus B	1969	470	Oriented	11,000[c]	6.0	**9.8**
ITOS (NOAA-1)	1970	393	Fixed Panels	10,260[c]	6.4	**8.7**
INTELSAT IV	1971	491[e]	Body Mounted	45,012[c]	3.0[e]	**2.24**
DSCS II	1971	520	Body Mounted	32,616[c]	4.52	10.4
SESP 71-2	1971	1550[d]	Oriented	34,500[c]	36.5	**8.8**

[a]The oriented systems, as used in this list, are considered to be any solar array system with some orientation capability.
[b]Including substrate (the base material upon which the silicon wafer is mounted).
[c]2 x 2 cm cells.
[d]Synchronous orbit
[e]Average

approximately 175 miles is not considered. Furthermore, they use oriented solar arrays only when the craft requires large amounts of power (more than 500 watts) and/or vehicle configuration and attitude control modes prevent acceptable performance from other systems.

With increasing demands for solar cell-generated power in spacecraft applications, various attempts are being made to increase the power-to-weight ratio and decrease the cost per unit power output. Current solar cell improvement efforts are primarily to optimize silicon solar cells to the 15 to 17 percent efficiency range. Engineers are doing some work on the development of GaAlAs/GaAs solar cells since they may be able to achieve 20 percent efficiency. However these cells will be a factor of 5 to 10 more expensive than silicon cells. Methods to reduce cost of solar cell power systems in the solar cell area include three areas, namely, (1) standardized on cell types and test procedures through development and use of a joint NASA-DOD specification, (2) study concentration concepts exploiting the higher temperature capability of the GaAs solar cells (the tradeoff here is increased area versus lower cost) and (3) observe the current solar power developments for terrestrial applications and exploit the low-cost, large-area technology for satellite application. Terrestrial solar array technology is working toward solar panels at less than one dollar per watt compared to 400 to 800 dollars per watt for current space technology.

Emphasis is being placed on improved packaging and deployment techniques. Proposed configurations include rigid, lightweight, biconvex solar array panels and various unfurlable and retractable systems. Several proposed systems function much like a window shade, which the craft can unfurl mechanically after being placed in its operational orbit, or trajectory. This type of array could be retractable for periods of powered flight such as during course correction and docking maneuvers.

Projected power densities using these various improvements are approximately 30 to 50 watts per pound; which is almost an order of magnitude improvement of current systems (see table 4-1). Solar cell proponents, who once thought their horizon limitations to be power levels of less than one kilowatt, are talking of arrays that can supply tens of kilowatts.

Solar thermal. Another possible way to utilize solar radiation is to develop solar heated systems using turboelectric, thermoelectric, or thermionic devices to convert focused solar thermal energy into electricity. The heart of the solar thermal systems is the solar mirror concentrator. Five-foot paraboloidal mirrors have been made that can focus over 85 percent of the incident solar radiation

through an aperature of 0.5 inch diameter. This would provide approximately two kilowatts of heat on a thermionic converter at 2,000° Kelvin. Increased power can be obtained by using solar concentrators that are composed of many petallike segments or that expand like umbrellas. This construction enables the creation of diameters up to approximately 50 feet with correspondingly higher power output.

Although estimates indicate that weight and cost advantages over other power systems may be possible, this is yet to be verified. In recent years, the government has either cancelled or cut back severely all major solar thermal programs. However, some interest still remains, since this type system would be more resistant to charged particle radiation and could operate closer to the sun than solar cell arrays. Conventional silicon solar cells do not function properly at temperatures in excess of 150° C.

Solar concentrator systems, coupled with thermal energy storage systems and heat pipes, are being considered for the spacecraft whose energy requirement is for thermal rather than electrical power. This is in addition to the solar concentrator systems' use as a heat source for conversion to electrical power. The energy storage system employs the freezing and melting of high-power density (100 to 200 watt hours/pounds) salts for thermal energy storage, and it uses high-temperature sodium heat pipes to transfer the heat from the concentrator through the energy storage system to the using spacecraft subsystem. During the sunlit portion of a satellite's orbit, an optical concentrator focuses sunlight on the cannister, melting the solid inside it. During shaded portions of the orbit, the molten material freezes, giving its heat to the temperature-sensitive components to maintain a proper temperature balance. Aluminum and alkali fluorides have been studied as possible heat absorbing solids, and Inconel cannisters as containers for the solid.

Nuclear Systems

The Atomic Energy Commission (AEC) and its successor, the Energy Research and Development Administration (ERDA), developed nuclear electric power systems under the Systems for Nuclear Auxiliary Power (SNAP) program. The objective of the program was to develop compact, lightweight, and reliable energy units and electrical power conversion devices for space, sea, and land uses. Since the Air Force was interested in long-life, lightweight, electrical power generation systems for use in space, it requested the AEC to develop such systems. The AEC initiated the SNAP program in 1955. During the lifetime of the SNAP program, it served both the Air Force and NASA. Although no longer in operation, the SNAP program was successful in reaching most of its objectives.

Under the SNAP program, ERDA developed radioiosotope and nuclear reactor systems. The odd-numbered SNAP, such as SNAP-3, used a radioiostope as an energy source, and the even-numbered SNAP, such as SNAP-8, used a nuclear reactor as an energy source. We describe some of the radioiosotope and reactor systems developed for space applications in the following sections.

Radiosotope. The basis for the radioisotope concept for utilizing nuclear energy to provide space power is the phenomenon of radioactivity, the spontaneous decay of unstable atoms. Nuclei emission of electrically-charged alpha particles, beta particles, and gamma rays accompanies this radioactive decay. The alpha particles are positive and are the same as the nucleus of the normal helium atom, two protons and two neutrons. The beta particles are negative and are identical to electrons. The designators, alpha and beta, are terms left over from the period when it had been determined that there were radioactive emanations, but their atomic identity had not been determined. Gamma rays, which often accompany radioactive changes, are high-energy penetrating electromagnetic radiation. Like high-energy X-rays, gamma rays can be very harmful to biological tissues. In large quantities, gamma rays will damage structural materials as well.

The half-life of the particular radioactive material expresses the rate of the radioactive decay process. It is a characteristic of each radioisotope species and is independent of the quantity of material or the time at which the decay measurements are started. Thus, the half-life of a radioisotope is that time at which its activity has decreased to one half of its original value beginning with any given starting time.

Radioisotopes, sealed in a container, generate heat when the container absorbes the radiation emitted by the radioisotope. A suitable conversion system converts part of the heat into electricity. Radioisotopes are a reliable source of heat, but they are dangerous to handle. If they emit gamma radiation, they require shielding. The lifetime designed into the power system depends on the half-life of the radioisotope. Providing an excess of radioactive material at the start, and just enough at the end of the operating period, can make compensation for the decay of the radioisotope. In the space environment, the spacecraft must eject by radiation the excess heat generated during the early life of the system. Figure 4-2 shows the variation in power output for polonium-210 (Po-210) with a 138-day half-life, curium-242 (Cm-242) with a 162-day half-life, cerium-144 (Ce-144) with a 285-day half-life, and promethium-147 (Pm-147) with a 2.6 year half-life.

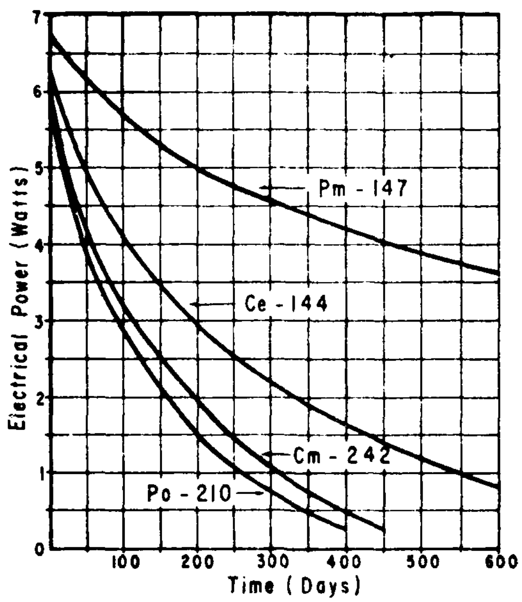


Figure 4-2 Isotope electrical power versus time

Although not shown in figure 4-2, the power output of a plutonium-238 (Pu-238) radioisotope system would be virtually a constant for the indicated 600 days. This is because the half-life of Pu-238 is 86.4 years.

The former AEC started the SNAP radioisotope power programs several years before Sputnik I. Although the radioisotope power programs suffered a setback in 1959 with the cancellation of an overambitious SNAP-1 and SNAP-1A program, isotopic power had developed to the point where it could not be neglected as a potential space power source. For example, the first proof of the practicality of the radioisotope power generation took place in January 1959, when researchers tested and delivered to the AEC the 2.5 watt SNAP-3, a polonium-210-fueled radioisotope generator. Polonium-210, which was suitable and readily available, is an alpha emitter. Figure 4-3 shows a cutaway diagram of this proof-of-principle device, which President Eisenhower introduced on 16 January 1959, as the SNAP-3 "atomic battery." It contained 27 lead-telluride thermocouple elements, treated with bismuth to produce negative and positive semiconductors. The thermocouple elements provided parallel paths for the heat to flow from the radioisotope source to the container radiator. The developers insulated these elements electrically from each other at both the hot (1,050°F) and cold (300°F) junctions. The complete SNAP-3 weighed only four pounds, and it was only 5.5 inches high and 4.75 inches in diameter.

Because of these efforts, it was possible with only a five-month lead-time to fabricate, fuel, test, and get approval to use a plutonium-238 fueled SNAP-3 radioisotope generator on the Navy's Transit-

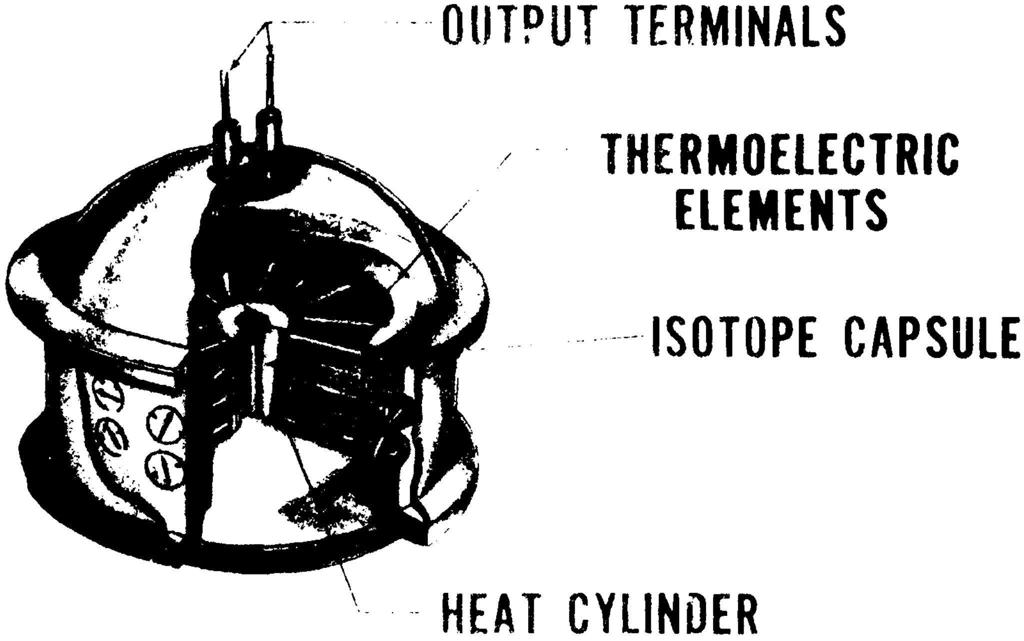


Figure 4-[illegible] SNAP-3 cutaway diagram

4-A navigational satellite, which they launched in June 1961. This was the first atomic power device used in space. The Pu-238 fueled SNAP-3 was a low powered (2.7 watts), lightweight, and rugged thermoelectric generator. The designers used plutonium 238 because of its extremely long half-life (86.4 years), its low gamma radiation, and its high thermal power density per unit of mass of fuel. The device weighed about 4.6 pounds. They designed it for a five-year life. It provided power for two of the experimental Doppler frequency navigational signal transmitters in the satellite.

Table 4-2 shows the highlights, performance, and status of the PU-238 fueled SNAP-3 radioisotope generators and additional space radioisotope power systems. Note that most radioisotope generators have power densities of approximately one watt per pound. The SNAP-27, a vital cog in the NASA Apollo Lunar Surface Experiment Package (ALSEP), showed the best power density performance, approximately 30 percent higher than the other SNAP devices. As far as the future is concerned, it appears that radioisotope generators can be built with power densities between two and three watts per pound. Although the power outputs of future systems probably will continue to be relatively low (less than 500 watts), higher power outputs are achievable.

Reactor. Nuclear reactor systems are considered to be the most attractive means of obtaining large quantities of space electrical power. Although at present there is no pursuit of the development of nuclear reactor space power systems, some have been developed already. Resumption of the space reactor program could occur if the need for very large quantities of electrical power becomes a reality.

A space nuclear power supply consists of three major subsystems: a nuclear reactor heat source and associated radiation shield, [illegible] subsystem, and a heat rejection subsystem. Such a system derives its energy from the controlled fissioning of uranium 235 in the core of the reactor. The fission energy [illegible] in the form of heat can be converted to electrical power by using one of three approaches: turboelectric, thermoelectric, or thermionic conversion.

As part of the initial SNAP program, the former AEC started the development of space reactor systems for generating electrical power [illegible] the development of a system

Table 4-2.

Space Radioisotope Power Systems

Designation	*Mission*	*Power (Watts)*	*Weight (lbs)*	*Isotope*	*Design Life*	*Status*
SNAP-1	Air Force Satellite	500	600	Ce-144	60 days	Program canceled in 1958.
SNAP-1A	Air Force Satellite	125	175	Ce-144	1 year	Program canceled in 1959. One unit successfully ground tested.
SNAP-3	Navy Navigational Satellites	2.7	4.6	Pu-238	5 years	2 in space, 6/61 & 11/61. First failed in late 60s. Second failed after 8 months.
SNAP-9A	Navy Navigational Satellites	25	27	Pu-238	6 years	3 launches; 9/63, 12/63, 4/64. Third launch failed to achieve orbit. Program now terminated.
SNAP-11	Moon Probe (Surveyor)	25	30	Cm-242	120 days	Mission canceled. One unit successfully ground tested in 1965.
SNAP-17	Communications Satellite	25	28	Sr-90	5 years	Design and test phase completed 12/64. Program now terminated.
SNAP-19 Nimbus	Weather Satellite	30	30	Pu-238	5 years	Initial launch 5/68 failed to achieve orbit. Two units successfully launched 5/69.
SNAP-19 Pioneer	Planetary Exploration (Fly-by)	30	30	Pu-238	5 years	Successfully launched 3/72.
SNAP-19 Viking	Mars Lander	[illegible]	[illegible]	Pu-238	5 years	Successful 1975 launch
SNAP-27	Apollo Lunar surface equipment package	60	46	Pu-238	2 years storage, 1 year lunar operation	Five units successfully deployed by Apollo astronauts
SNAP-29	Short lived DOD and NASA earth missions	400	[illegible]	Po-210	90 days	Program terminated 6/69.
Multi Hundred Watt (MHW)	Lincoln Laboratories experimental satellite	150	80	Pu-238	5 years	Successful 1976 launch

for an Air Force advanced satellite. In the spring of 1956, the Air Force selected Atomics International's concept, using a mercury-Rankine cycle, and designated it SNAP-2 (fig. 4-4). The objective of this program was to develop a nuclear turboelectric unit generating three kilowatts for use in space. The SNAP-2 reactor uses 37 cylindrical fuel rods consisting of uranium-235 intimately mixed with zirconium hydride, which serves as a moderator to slow down the neutrons that maintain the fission chain.

A sodium-potassium alloy, in the liquid metal state, transfers the heat to the conversion equipment and cools the reactor by circulating through the reactor core and a mercury boiler-heat exchanger. The coolant enters the core at a temperature of approximately 1,000°F and leaves at about 1,200°F. The exchange transfers heat from the liquid sodium potassium alloy to liquid mercury, the working fluid, converting the mercury into superheated vapor. The mercury vapor expands through a turbine to drive an electric generator so that it produces electricity. To complete the cycle, the mercury vapor leaving the turbine condenses to liquid mercury in a radiator before it returns

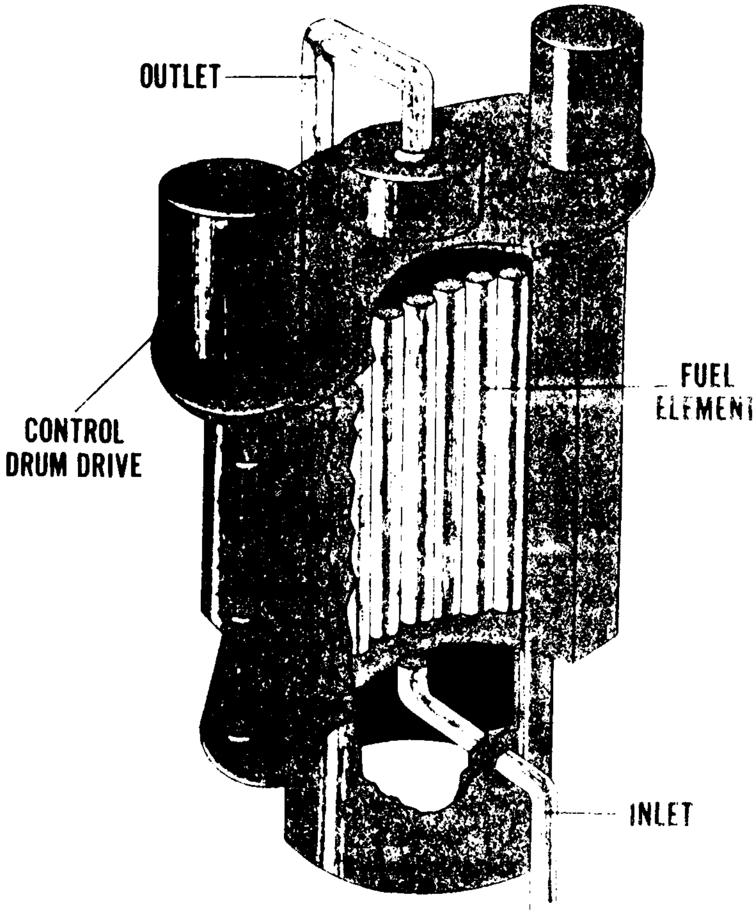


Figure 4-4. SNAP-2 reactor

to the boiler. The estimated weight of the SNAP-2, when installed in a space vehicle as a "nose cone module" and including sufficient shielding for protecting transistors for one year, was 900 pounds.

Several experimental and developmental reactors were designed, constructed, and operated. They produced approximately 500,000 Kw_t hr (thermal power) from 1959 through 1962 without failure. In 1963, the government terminated the SNAP-2 program as a system, but the AEC continued a base program to develop and demonstrate the potential of the uranium-zirconium hydride fuel technology and to improve the power capabilities of the mercury-Rankine conversion system.

The first reactor-powered electrical system flight tested in space was the SNAP 10A. The SNAP-10A derived its thermal energy from a power output modified SNAP 2 reactor. In the SNAP-2, the reactor produces about 50 Kw_t at a 1,200°F outlet temperature. The SNAP-10 reactor produces about 30 kw_t at an outlet temperature of 990°F and an inlet temperature of 880°F.

Furthermore, instead of using a turboelectric heat conversion system, a silicon-germanium thermoelectric device converted the heat from the SNAP 10A nuclear reactor directly into electricity. This device formed an integral part of the radiator (see fig. 4-[illegible]).

An Atlas Agena vehicle launched the SNAP 10A into an earth orbit in April 1965. The system was designed to provide 500 watts of electrical power at 28 volts of direct current for a period of one year in a space environment. The power system obtained full power operation approximately 12 hours after liftoff and operated successfully and continuously at full power for 43 days. From

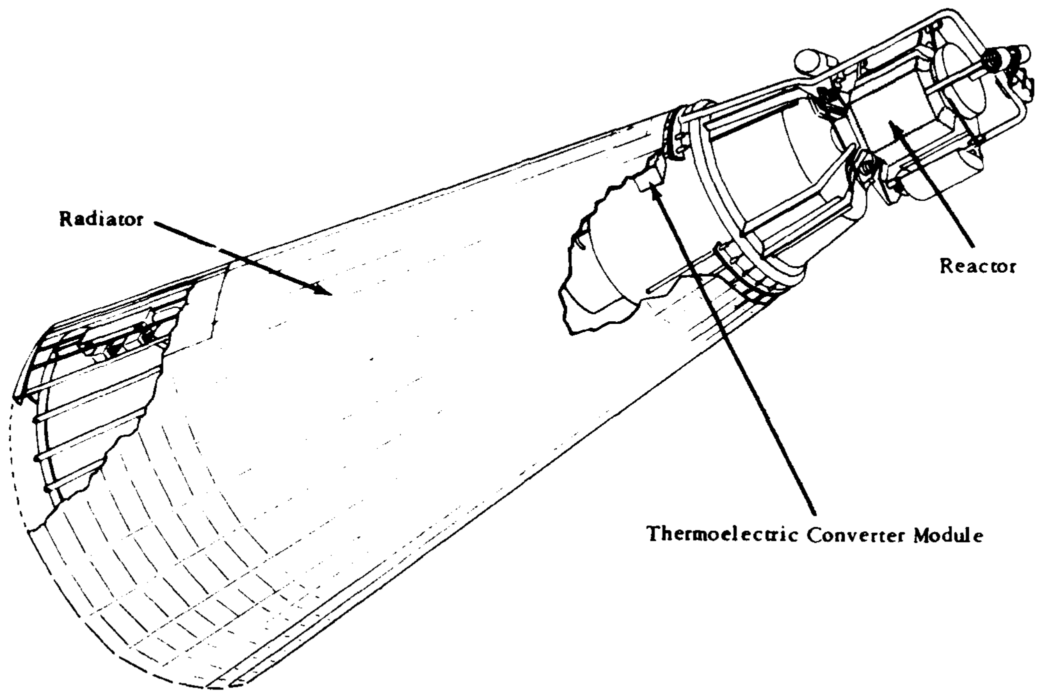


Figure 4-5. SNAP-10A system.

telemetry data, NASA established that the probable cause of the shutdown was a voltage regulator failure. This resulted in high voltage that caused other components to fail. During the 43 days of successful opeation, the SNAP-10A produced 500,000 watt-hours of electricity, and it met all of its objectives with the exception of the endurance demonstration.

Safety

No discussion of the use of nuclear electric power systems in space flight vehicles would be complete without considering potential radiological hazards. Conceivably, these devices could present hazards during prelaunch, during powered flight to orbit or to escape velocity, during orbit, or during reentry. Therefore we must consider the following three basic safety objectives in the design of these devices: (1) under the worst conditions, these devices should not materially increase the general atmospheric background radiation; (2) at the launch pad, harmful radiation should not extend beyond the device itself or the normal exclusion area; and (3) the device should not produce a local hazard upon return to earth.

Safety has been the basis of many design decisions in the development of SNAP. Compromises in the design of a system have been necessary to achieve a suitable balance between safety and operation characteristics such as weight, simplicity, and reliability. The design decisions have resulted in encapsulated radioisotope heat sources that can withstand the very high overpressures and temperatures that could occur if there were an explosion or fire on the launch pad. Reactor systems are being designed so that full-power reactor operation does not begin until they have placed the device safely in orbit. Reactor systems and components have been designed to ensure burnup at high altitude during reentry. Radiological hazards should not significantly limit the use of nuclear electric power systems in space, when these hazards are anticipated and appropriate measures are taken in design, in establishing handling procedures, and in fixing limits for operation.

SELECTION OF POWER SYSTEMS

The selection of a power system is primarily a function of the space mission and its power requirements in magnitude and duration. However, there is a host of other factors, including economic considerations, size and weight, orbital altitude, reliability, and requirements for attitude control. All of these factors make the selection and design of a power system for each specific space mission a custom built affair.

While many types of space power systems are possible, the primary system has been the solar photovoltaic/secondary battery system. Future systems are likely to continue utilizing this combination with advanced components such as lightweight metal-gas batteries, higher efficiency solar cells, and increased life for all power system components.

Although radioisotope systems operate in the same power levels and mission duration times as solar cell/battery systems, their potential seems to be limited to special missions where unique radiation hardening, deployment, or energy requirements may be necessary. We can only consider nuclear reactor systems when we require a very high-level power, and even then they must compete with solar concentrator/thermal energy systems for these as yet unspecified applications. For the short duration power requirements, primary batteries will continue to dominate although we might use fuel cells for certain specific applications.

BIBLIOGRAPHY

"Advanced Space Reactor Concept." *Space World*, November 1970.

Austin, L.G., ed. *Fuel Cells, A Review of Government-Sponsored Research 1950–1964* (NASA-SP-120). Washington D.C.: Scientific and Technical Information Division, Office of Technology Utilization, NASA, 1967.

Boretz, J.E. "Power Systems," *Space/Aeronautics*, Research and Development issue (31 July 1968):85–90.

Corliss, Williams R. and Robert L. Mead. "Power from Radioisotopes," *Space World* (May 1970):41–45.

Crossley, P.A. *Review and Evaluation of Past Solar Cell Development Efforts* (NASA-CR-86979). Princeton, N.J.: David Sarnoff Research Center, Astro-Electronics Applied Research Laboratory, 1967.

Francis, Howard T. *Space Batteries* (NASA-SP-5004). NASA, 1964.

Glasstone, Samuel. *Sourcebook on the Space Sciences*. Princeton, N.J.: Van Nostrand Co., 1965.

Gylfe, J.D. and R.E. Wimmer. *Reactor-Thermoelectric Power Systems for Unmanned Satellite Applications*. Technical Paper, Intersociety Energy Conversion Conference, Los Angeles, Calif. September 1966.

Haloey, D.S., Jr. *The Coming Age of Solar Energy*. New York: Harper and Row, 1963.

Hanchett, James. "Whither Now, SNAP-8," *Necleonics*. Vol. 25, no. 5 (May 1967):54–57.

Hough, W. W. *The Effect of Launch Time on the Performance of a Solar Array/Battery Electrical Power System* (NASA-CR-88404). Washington D.C.: Bellcom, Inc., 1967.

"Huge Solar Array for Space Station." *Space World* (March 1971):50.

"Ion Engine to be Tested in Orbit." *Space World* (June 1969):7.

Klass, P.J. "Gains Cited in Thin Film Solar Cell Efforts." *Aviation Week and Space Technology* (26 August 1968).

Mackay, Donald B. *Design of Space Power Plants*.Englewood Cliffs, N.J.: Prentice-Hall, 1963.

"New Modular Space Power System Under Study." *Space World* (July 1971): 36–37.

"New Windowshade (of Solar Cells) to Harness Sunlight for Satellite Power," *Space World* (December 1970):38.

"Nuclear Energy in Space and the Grand Tour," *Chemistry* (1 February 1969):20.

"Providing Perpetual Power," *Space World* (December 1970):36–44.

Scientific and Technical Division, NASA. "Space Power Systems Advanced Technology Conference." (NASA-SP-131). Washington D.C.: Scientific and Technical Division of NASA, 1966.

Snyder, Nathan W., ed. "Energy Conversion for Space Power." Paper presented at ARS Space Power Conference, Santa Monica, Calif. 1960. Published in *Progress in Astronautics and Rocketry*, vol. 3. New York: Academic Press, 1961.

——————. "Space Power Systems." Paper presented at ARS Space Power Conference, Santa Monica, Calif. 1960. Published in *Progress in Astronautics and Rocketry*. Vol. 4. New York: Academic Press, 1961.

"Spaceship Power 'On Full'." *Space World* (July 1971):12–15.

"Space Station: Key to the Future." *Space World* (February 1971):6–44.

Szego, George C. and J. Edward Taylor, eds. "Space Power Systems Engineering." Paper presented at American Institute of Aeronautics and Astronautics Aerospace Power Systems Conference, Philadelphia, Pa. 1964. Published in *Progress in Astronautics and Aeronautics*. Vol. 16. New York: Academic Press, 1966.

Zipkin, Morris A. and Russell N. Edwards, eds. "Power Systems for Space Flight." Paper presented at American Rocket Society Space Conference, Santa Monica, Calif. 1962. Published in *Progress in Astronautics and Aeronautics*, vol. 11, New York: Academic Press, 1963.

Chapter 5

GUIDANCE AND CONTROL

The three processes of ICBM warhead delivery on target, satellite placement in and maintenance on orbit, and the recovery of a vehicle from space have one thing in common. Some form of guidance, navigation, and control (GNC) is required.

Navigation is the determination of the current state of motion of the vehicle, that is, position, velocity, and attitude. We determine this state with reference to some mission-dependent coordinates suitable for defining the motion of the vehicle. Guidance is the computation of corrective actions to change from the navigation-determined vehicle state to a required vehicle state. The required state will depend on particular mission requirements. Mission requirements may specify a required present state, for example, a particular orbit; or an objective, for example, rendezvous at a point in space from which we can compute the present required state. Control is the application of corrective maneuvers to obtain the changes commanded by guidance.

We might define three broad categories of space missions as ground-to-ground (missile) missions, earth satellite missions, and deep space missions. Since we have conducted deep space missions to gain basic knowledge pertaining to our universe, and they are not directly of a military nature as yet, we will not discuss their requirements for GNC. The matrix of figure 5-1 sets forth the main GNC functions for the various phases of the first two mission types mentioned previously. The following sections contain an expanded discussion of two representative functions from this matrix. However, we begin with a general discussion of some techniques used to perform guidance, navigation, and control.

GUIDANCE, NAVIGATION, AND CONTROL TECHNIQUES

One of the basic elements necessary for success in space, be it military or nonmilitary, is an infallible navigational system. You can imagine the inoperability of any space mission without it. This section discusses the various techniques used in successful space navigation.

Navigation

Navigation computes the present state of the vehicle from measured values of selected physical quantities. This statement implicitly requires a known relationship between each particular measured quantity and the variables used to define the vehicle state. Frequently, the state-defining variables are not measured directly. Navigation measurements include optical sightings, phase shifts and time delays of radio waves, reaction of inertial sensors to motion, and combinations of these.

Optical sightings provide angle information as measured between the line of sight of an optical instrument and some other established direction in space. The other direction may be another optical line of sight, or a direction established by orientation of an inertially-stabilized platform, among others. We may compute position and attitude from such angle information.

Radio navigation, measuring radio-signal phase shifts and time delays, can provide information on range, range rate, and direction of the line of sight between the transmitters and vehicle. We can

	NAVIGATION	GUIDANCE	CONTROL
Boost	Compute present position, velocity, and attitude	Use precomputed early boost trajectory to pass through region of high dynamic pressure. During late boost compute velocity corrections necessary to place vehicle on desired trajectory or orbit.	Maintain vehicle attitude and attitude rates within safe limits during region of high dynamic pressure. Direct thrust vector to provide velocity changes commanded by guidance.
Orbit	same	Compute deviations from stable attitude or required attitude change. Compute deviations from required orbit or changes needed to obtain new orbit.	Change attitude via thrusters, reaction wheels, electromagnetic torques. Stabilize via spin stabilization, control moment gyro, reaction wheels, etc. *Use passive stabilization* from solar pressure, gravity gradients. Thrust to new orbit.
Recovery or Rendevous	same	Compute deviations from re-entry attitude. Compute desired direction and magnitude of available aerodynamic forces to shape reentry trajectory. Compute orbit changes needed to achieve rendezvous. Compute range, range rate, and attitude during final stages of rendezvous.	Change attitude via thrusters, aerodynamic control surface, C. G. shift. Thrust to new orbit. Thrust and change attitude as necessary to finalize rendezvous.

Figure 5-1. Guidance, navigation, and control functions versus mission phases.

locate the transmitters at precisely-known ground locations or in orbit, for example, the NAVSTAR, a global positioning system. We can combine these measurements to provide computed position and velocity.

The two sensors used, gyroscopes and accelerometers, describe inertial navigation. Inertial navigation gyroscopes detect rotation about one or two input axes, the first case being a single-degree-of-freedom and the second, a two-degree-of-freedom gyro. By suitable mechanization, three single-degree-of-freedom or two two-degree-of-freedom gyros provide measurements necessary to define a coordinate frame fixed to the platform on which the gyros are mounted. The platform coordinate frame is defined in the sense that we can compute navigation state variables from accelerometer measurements taken in platform coordinates. The accelerometers mounted on the platform are calibrated to measure three components of specific force. Then, specific force is the acceleration of the platform due to all forces except gravity and any other field forces that are usually negligible with respect to gravity (see fig. 5-2).

By measuring specific force and computing analytically a value for gravity, we can calculate the change of the state variables from their initial values. We add these changes to the initial values, which we must provide prior to starting an inertial navigation portion of a mission. Thus, the navigation consists of computing total acceleration as the sum of measured specific force and computed gravity, obtaining changes in velocity and position by integrating the acceleration, and obtaining updated velocity and position by adding to the initial-condition values. We can obtain vehicle attitude information from gimbal-angle readout devices, or we can compute it as part of the navigation process in strapped-down inertial systems.

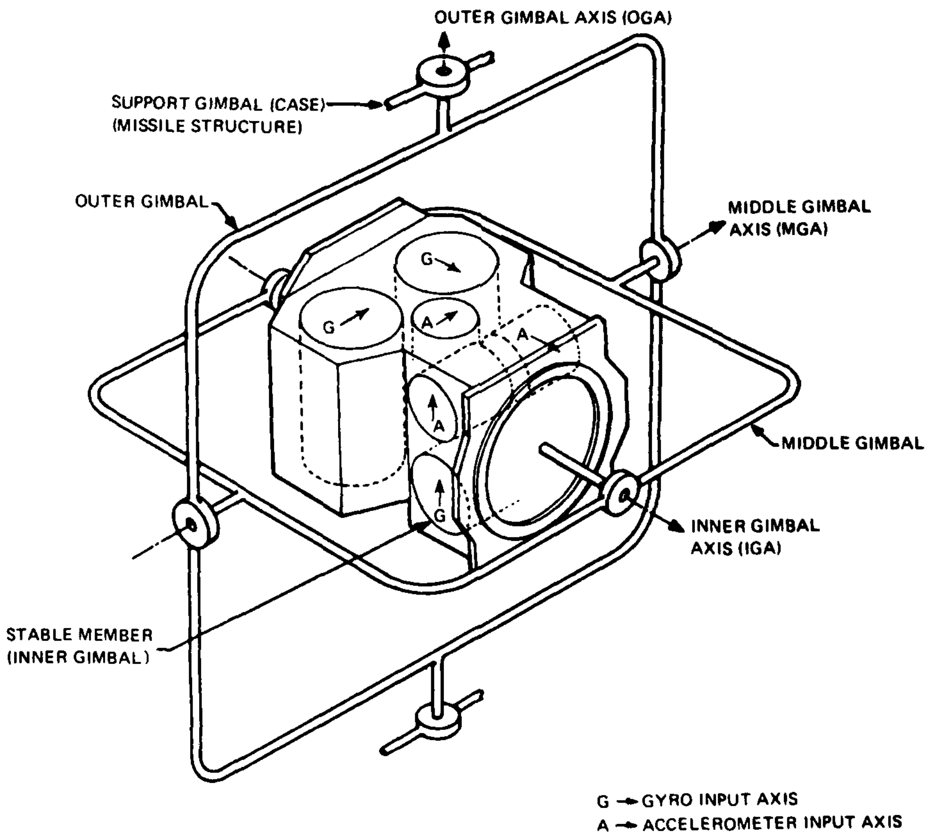


Figure 5-2. Simplified diagram of stable platforms.

We can combine the above types of navigation measurements, in a hybrid configuration, taking advantage of the best features of each. Of special importance in recent years is the optimal combining of various linearized measurements to yield a best estimate of the navigation-state variables. This is Kalman, or optimal, filtering.

Guidance

Guidance is the process of comparing the measured navigation state with a required one and then computing commands to correct differences between the two. The mission controllers might compute continuously the required navigation state based on mission objectives or on the present value of a navigation state predetermined as a function of time. For purposes of defining corrections, the state variables conveniently break out into two distinct types of motion: translational motion for the position and velocity variables and rotational motion for the attitude variables. Using this breakdown, the guidance computations yield required velocity corrections (translation) and required attitude changes (rotation). The velocity correction may be the so-called velocity to be gained, that is, the velocity increment presently necessary to have the vehicle pass through a given point in space at a later time. It may be the velocity correction necessary to get the vehicle back on a prescribed trajectory. Attitude change could be the rotation of the vehicle back to a desired stable alignment or the rotation from one attitude to another attitude, perhaps to look at a different star.

Control

Control is the application of forces to the vehicle to drive the state variables to required values. In discussing control, it is convenient to use the translation and rotation aspects of the vehicle motion as we did with guidance.

Basically, translation control is the directing of a control force to bring about a commanded velocity change. Two methods are available generally. For control outside the atmosphere, the spacecraft controls the thrust direction and burn time of a rocket motor to generate a desired velocity change. Less used is an additional control of the magnitude of the thrust. For motion inside the atmosphere, for example, during reentry, the possibility of using aerodynamic forces exists. By providing attitude control on a vehicle that can develop lift, mission control can orient the lift and drag vectors to bring about desired velocity changes.

Applying torques to the vehicle or using the conservation of angular momentum can accomplish attitude control. Two common ways to apply torques are use of reaction control jets and aerodynamic control surfaces. Less obvious torques acting on an orbiting vehicle are gravity gradient torques and electromagnetic torques due to interaction with the earth's magnetic field. In the frictionless environment of space, the very slight difference in gravity acting on portions of the satellite that are at different distances from the earth is sufficient to apply torques on the satellite. For a long satellite, the tendency is to have the long axis point to the earth. The satellite will oscillate about this direction unless some form of damping is used to reduce the oscillations to zero. Similarly, the interaction of electrical currents inside the satellite with the earth's magnetic field can cause the satellite to rotate in the frictionless environment.

The spacecraft can use conservation of angular momentum to control vehicle attitude. By increasing the angular velocity of an inertia reaction wheel inside the vehicle, the vehicle will rotate in the opposite direction to conserve the total angular momentum of the satellite. This is the rotational form of Newton's third law, which states that an action by the vehicle on the wheel produces an equal and opposite reaction from the wheel on the vehicle (assuming no external torques on the vehicle). Similarly, applying a torque to a gyro inside the vehicle can make the gyro precess. This precession is a change of angular momentum and requires, by Newton's third law, an opposite change in angular momentum, which comes about by the vehicle rotating appropriately. We should note that it is possible to saturate both devices such that they cannot accept any more input

angular momentum. When this happens, applying torques to the vehicle from other sources such as jets or electromagnetic interaction desaturates the device.

GUIDANCE, NAVIGATION, AND CONTROL FUNCTIONS

The first item from the matrix of figure 5-1 that we will discuss in expanded form is the boost or injection phase. We will use the GNC of a ballistic missile to focus on this mission phase.

Boost and Injection

In its simplest form, a modern ballistic missile weapon system has two major parts, namely, a booster rocket of one or more stages and a warhead housed in a protective structure of particular shape to survive a high-speed reentry passage back through the atmosphere. We recognize that there are three distinct phases of flight. The first is powered flight from launch, which accelerates the missile upward through the earth's atmosphere and down range toward the target, to rocket thrust termination at the warhead release point at high velocity above the atmosphere. The second is the ballistic coast of the warhead in a free-fall trajectory in the vacuum of space. The third is the reentry of the warhead back through the atmosphere.

The forces acting on the rocket determine the motion of the weapon through these three phases. The large propulsive force of the rocket accelerates the payload in a direction opposite to that of the exhausted gases of the burning propellant. During the atmospheric passage of the rocket boost and warhead reentry phases, large aerodynamic forces of drag and lift affect the motion. During all the phases, the force of gravity pulls the vehicle toward the center of the earth.

Only the rocket propulsive force is under direct control to affect the aim of the weapon. The accuracy of a ballistic missile depends upon the ability to steer the rocket thrust so that the position, speed, and direction of motion at thrust termination are precisely those to establish a trajectory that will hit the target. But the accuracy also depends upon the ability to predict the motions in the free-coasting trajectory and the ability to specify target location coordinates accurately.

The guidance of a ballistic missile booster rocket requires the accurate measurement of motions as they occur during rocket-powered flight. The process of measurement can utilize either radio or radar techniques to measure the motions directly or inertial-sensing techniques to measure forces causing the motion.

By using radar stations and radio-ranging equipment located on accurate baseline arrays, and interrogating cooperative transponder beacons on the missile, radiation sensing can provide almost micrometer precision in the measurement of location and velocity of the burning rocket. Some characteristics of these radiation-sensing methods are undesirable, especially for a weapon system. An obvious drawback is that the ground stations are complex, conspicuous, and, therefore, vulnerable to enemy countermeasure activity. For this reason, the guidance of modern ballistic missile rockets has depended upon, and caused remarkable advances in, the technology of inertial sensing. Inertial sensing is completely self-contained, having no active or passive radiation contact with the outside world.

Inertial Sensing

The principle of inertial sensing depends upon the direct measurement of force acting on a test mass inside an instrument known as an accelerometer. There are two prime forces acting on the accelerometer's test mass. One is the force of gravity. The other is the inertial-reaction force, primarily from the propulsion system. The latter causes an equal and opposite force on the test mass by a spring attached between the mass and the accelerometer case. As both of these forces have direction and magnitude, we must treat them analytically as vectors.

The vector force of gravity, $\bar{f}_g$, on the test mass is proportional to both the magnitude of the mass, m, and to the magnitude of the local acceleration of gravity, g, and has the direction of the local force of gravity:

$$\bar{f}_g = m\bar{g}. \tag{1}$$

Components within the instrument measure the vector inertial reaction force, $\overline{f}_a$. The vector sum of the gravity and spring forces is the total force on the test mass causing its motion and its proportional to the net acceleration of the test mass,

$$\overline{f}_a + \overline{f}_a = m\overline{a} \tag{2}$$

Solving for this acceleration and substituting equation 1 gives

$$\overline{a} = \frac{\overline{f}_a}{m} + \overline{g} \tag{3}$$

This is the fundamental equation of inertial navigation. $\overline{a}$ is the acceleration of the test mass, and therefore, is the acceleration of the missile. $\overline{g}$ is the local acceleration of gravity, which is accurately known as a function of the position of the vehicle. Considering only the earth's attraction,

$$\overline{g} = \frac{-GM_e\overline{r}}{} \tag{4}$$

where GM_e is the earth gravitational constant and $\overline{r}$ is the position vector of the missile relative to the earth's center.

The term $\overline{f}_a/m$ in equation 3 is the specific force (force per unit mass) measured by the accelerometer, and has units of acceleration. This specific force equals the acceleration due to directly applied forces. The directly applied forces are those from the rocket propulsion, aerodynamic lift and drag, and any other miscellaneous externally applied forces except that of gravity.

We must emphasize this important point of inertial guidance: the accelerometer cannot measure the component of acceleration due to the force of gravity. We must determine this part of motion analytically from the known magnitude and direction of gravity as a function of position.

From the basic equation of inertial navigation, equation 3, we can determine the total local acceleration as the vector sum of the accelerometer-measured specific force and the analytically-determined acceleration of gravity, such as given in equation 4, as a function of missile position. We can integrate this equation once to get velocity and position as embodied in the constants of integration.

At this point, some of the practical aspects of implementation need introducing. Equation 3 is a vector equation. Equipment to deal with this equation will operate on vector components represented in an appropriate reference coordinate frame. Suitable reference coordinate frames that make equation 3 true are nonrotating and nonaccelerated in the sense that the center of the earth is nonaccelerated, being in free-fall in its motion in the Solar System. For ballistic missile inertial guidance, it is common practice to implement the acceleration equation in an earth-centered nonrotating coordinate frame. It is a common practice to mount accelerometers on a gyroscopically-stabilized platform so that the accelerometer specific-force output signals are directly components of acceleration in a nonrotating coordinate frame.

By the action of conservation of angular momentum of their rapidly spinning wheels, gyroscopes can generate signals proportional to the angular displacement in space of their cases. If the gyros are mounted suitably on a stable platform, such as the three-degree-of-freedom, gimbal-supported member figure 5-2 represents, the gyro signals become attitude-error signals for servo electronics. These electronics can drive gimble-axis torque motors to keep the inner gimbal spatially nonrotating in spite of the rotations of the missile and stable platform support structure. It will hold an orientation in space established by an initial alignment process accomplished on the ground before launch. The angular rotations, measured by mechanical-to-electrical angle signal generators mounted on each axis of the gimbal system, are the components of missile angular motion with respect to the reference orientation determined by the prelaunch alignment. These are the rotational outputs, roll, pitch, and yaw, of the inertial sensing. The translational outputs are the components of specific force measured bv the accelerometers mounted on the nonrotating platform.

Gyroscopes and accelerometers are the basic instruments of inertial sensing. There have been many successful design approaches of these instruments, incorporating a variety of ingenious features, motivated towards optimizing various measures of performance. Figure 5-3 is a cutaway view of the single-axis gyroscope used in Apollo spacecraft guidance. (The term, single axis, refers to the fact that the gyro is sensitive to input angular motion only about a single axis.) The platform requires three of these, with the orthogonally-oriented input axes, for stabilization. Figure 5-4 shows a cutaway view of the single-axis pendulous accelerometer, also used in Apollo guidance. The test mass is configured as a single-axis pendulum, restrained by torque feedback to a reference angular position with respect to the case. The torque required is the measure of specific force. Three of these instruments with orthogonally-oriented input axes are necessary to measure all components of the specific force. Figure 5-5 shows implementation of a complete gimbal-supported stable-platform inertial sensor.

Inertial Navigation

The use of inertial sensing to guide a missile rocket can take a number of different forms. We have chosen a conceptually simple form for illustration. It involves several steps. The first step integrates the basic equation of inertial sensing to produce explicit signals representing missile translational velocity and position. We call this process inertial navigation, and illustrate it in figure 5-6. Since the accelerometers do not measure the components of acceleration due to gravity forces, these are computed separately and added to the accelerometer signals as shown. The integration of this net acceleration, the left side of equation 3, is the velocity change. The addition of this change to the initial velocity at launch results in the indicated-velocity output signal. The indicated position change (provided by the integration of the indicated velocity) added to the initial position gives the net indicated position signal from inertial navigation. Since local-gravity acceleration is a known function of missile position as equation 4 gives, then the indicated position is the necessary input to the gravity-acceleration computer. All variables are vectors, but the components perform the actual implementation. There are three outputs of this inertial navigator, namely, indicated thrust acceleration, indicated velocity, and indicated position. The inertial guidance function described in the next section uses all three.

Inertial Guidance

The function of inertial guidance is to steer the rocket thrust to achieve the desired terminal condition. During passage of a ballistic missile rocket up through the atmosphere, the objective of steering is to control the motion to pass safely through the period of high-aerodynamic loading. During this period, the missile can tolerate only a small angle of attack and still maintain control and structural integrity. It is usual to steer this early phase through an angle history that is a predetermined function of time from launch. Once safely out of the atmosphere, the guidance steering is motivated to reach computed rocket-cutoff conditions.

Figure 5-7 represents a block diagram for one concept of ballistic missile inertial guidance. Note that we have abbreviated the previous figure for inertial navigation and represented it by a box on the left of figure 5-7. Guidance must steer the rocket so that it will release the warhead on one of the many trajectories that will hit the target. The input to the box labeled "required velocity computer" specifies the target to the system. The function of this box is to compute the required velocity vector the missile should have at the present position and instant of time so that the warhead, if released to coast free now, would be on a trajectory having the desired properties. The difference between the required velocity and the present indicated velocity is the velocity the rocket will gain. If the rocket's steering is proper, the rocket will reduce the gained velocity towards zero. When it achieves zero, the missile control should cut off the rocket engine thrust as shown by sending the thrust-termination signal. The steering strategy illustrated aims the missile so that the direction of the present measured thrust acceleration is parallel to the direction of the velocity that the rocket will gain, thus assuring that all components of the velocity that the rocket will gain will be reduced to

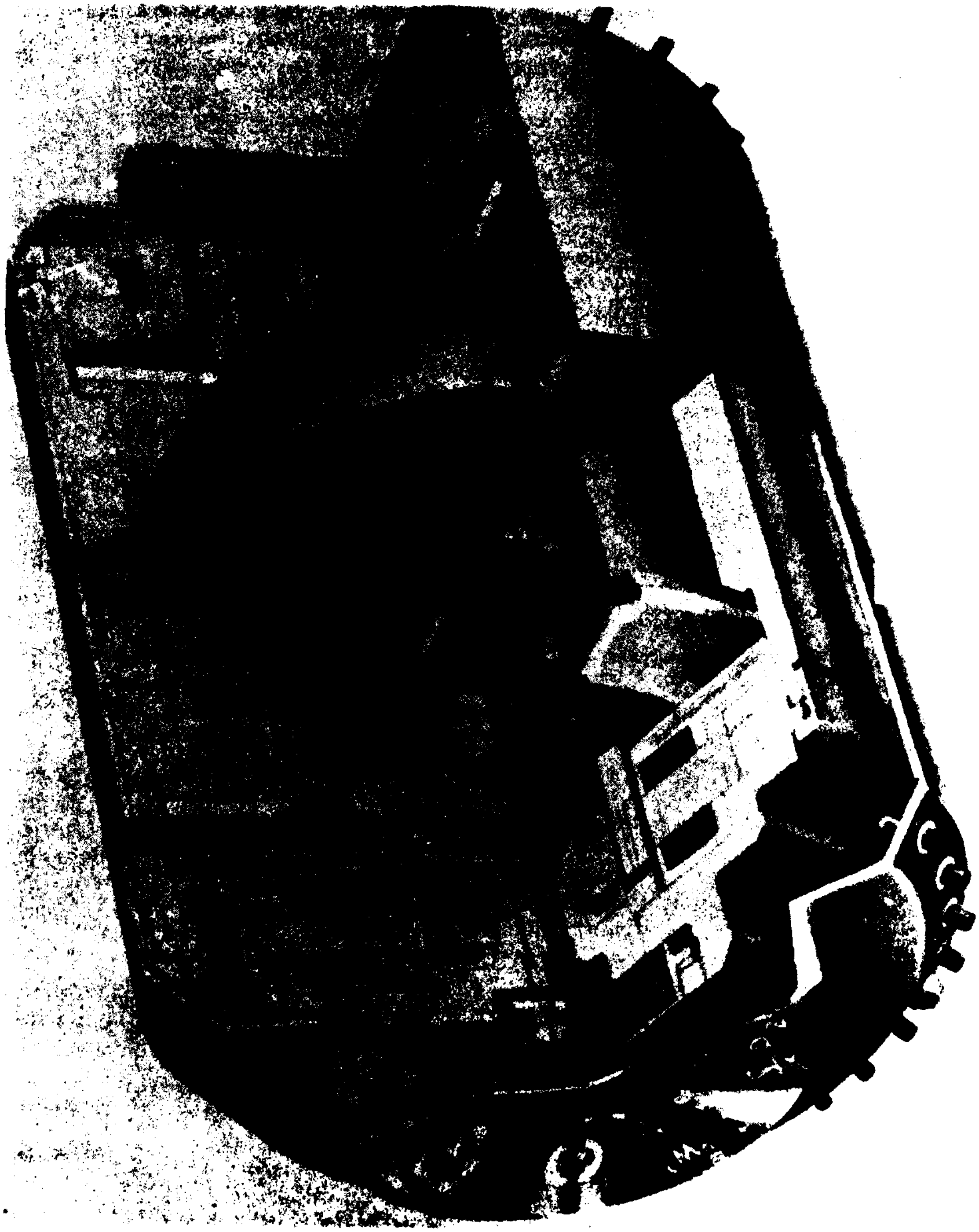


Figure 5-3. Single-axis gyroscope.

Figure 5-4. Single-axis pendulous accelerometer.

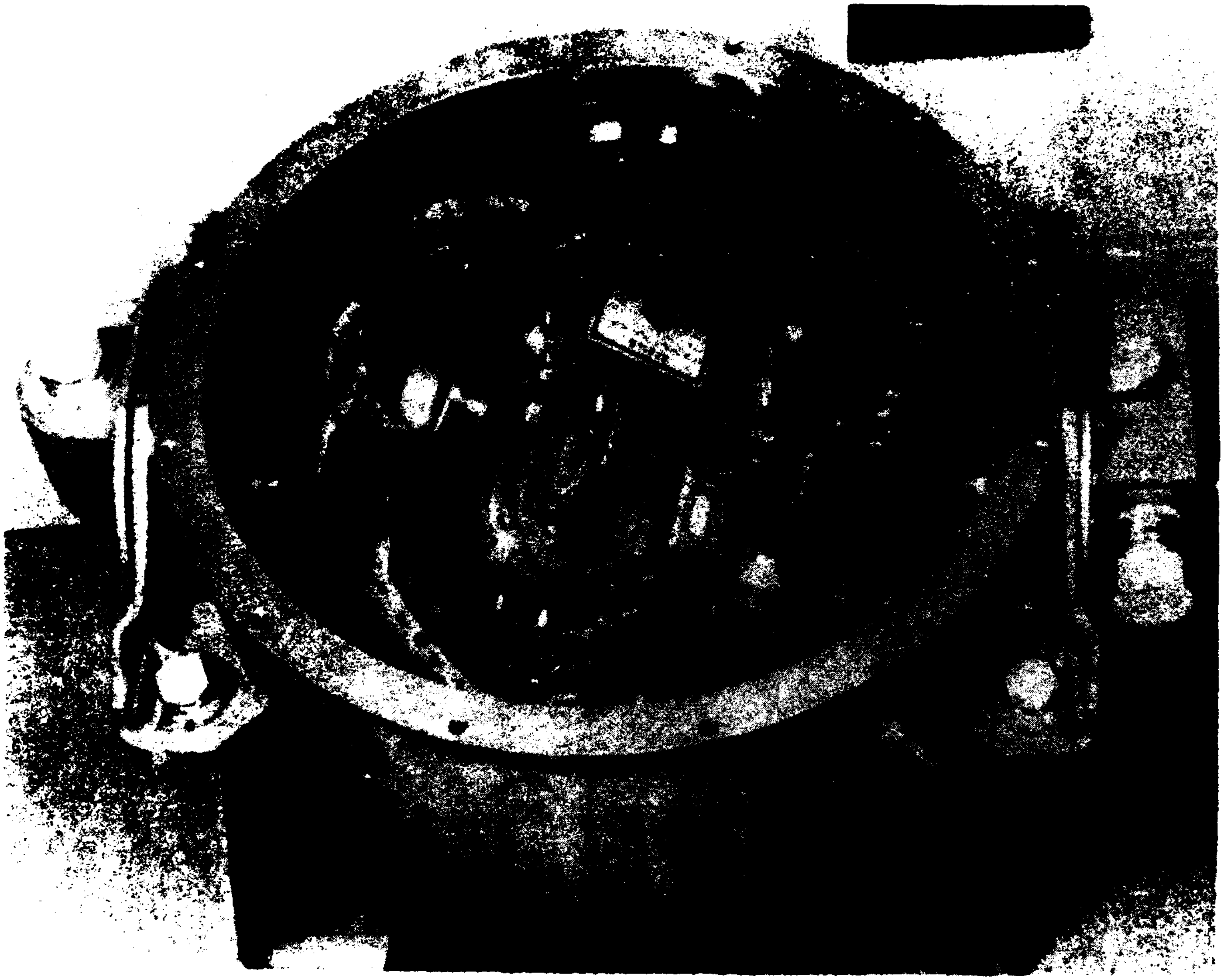


Figure 5-5. Stable platform.

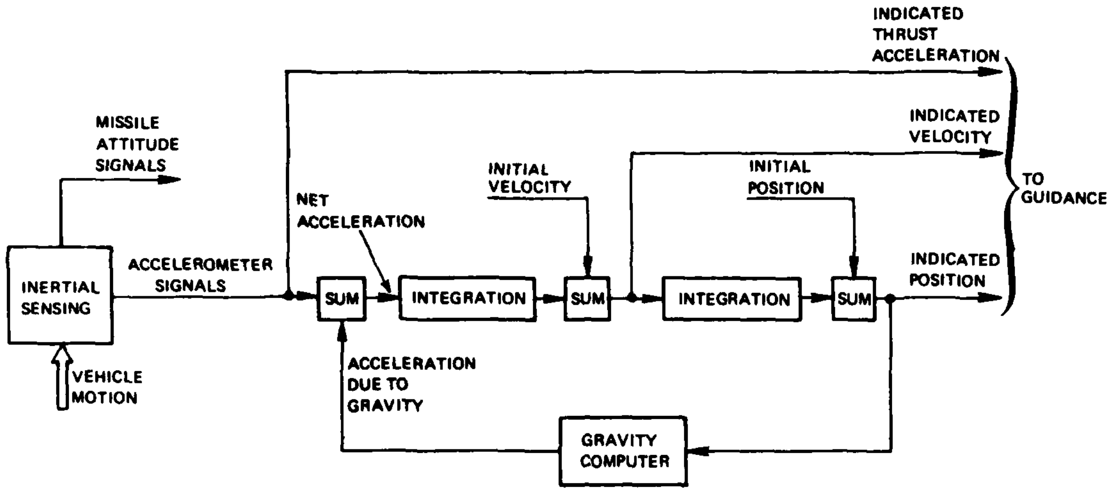


Figure 5-6. Inertial navigation.

zero simultaneously as desired. The steering computer causes the thrust acceleration to be parallel to the velocity that the rocket will gain by creating an angular-velocity command vector that will cause the missile to rotate the thrust vector towards the velocity-to-be-gained vector. This angular-velocity command goes to the missile's autopilot or thrust vector control system described in the following section. This is one representation of a particular guidance scheme called velocity-to-be-gained steering.

This seems an appropriate place to comment on the powerful implementations of the simple event of the guidance system thrust termination signal. A ballistic missile, with its high-performance rocket stages and associated equipment, depends upon many complex systems working successfully to complete its mission. If the guidance system senses that all three components of the velocity that the rocket will gain have passed sufficiently close to zero simultaneously so as to generate the thrust-termination signal, then there should be almost no failure anywhere in the whole missile system up to that time that will affect the performance of the weapon significantly. The only outright failure that would mislead such interpretation is the logic and circuitry generating the thrust-termination signal itself. Once signal generation occurs, we are assured that everything else worked as intended, including most of the guidance system as well. The event of three components of velocity arriving at zero simultaneously with a significant failure in achieving the velocity components correctly or in measuring them properly is nearly impossible. Large degradation in performance, not outright failure, in the inertial sensing instruments could fool this criterion of indicated success, but experience with these instruments and simple analysis show that such a situation is most rare compared to the many other possible failure modes elsewhere in the missile system. If the thrust termination signal occurs and the thrust actually ceases as sensed by the accelerometers, then the system has sent the weapon successfully on its intended free-fall ballistic trajectory. We can use such a signal to arm the warhead with the assurance that there is negligible chance that it will go awry and explode far from the intended target. If there is no generated signal, the missile could send a simple one-bit coded signal to the ground to indicate failure and the need for another launch against that target.

Thrust Vector Control

Thrust vector control is the closed-loop process that keeps the vehicle from tumbling under the high forces of engine thrust and accepts turning or guidance steering commands to change the

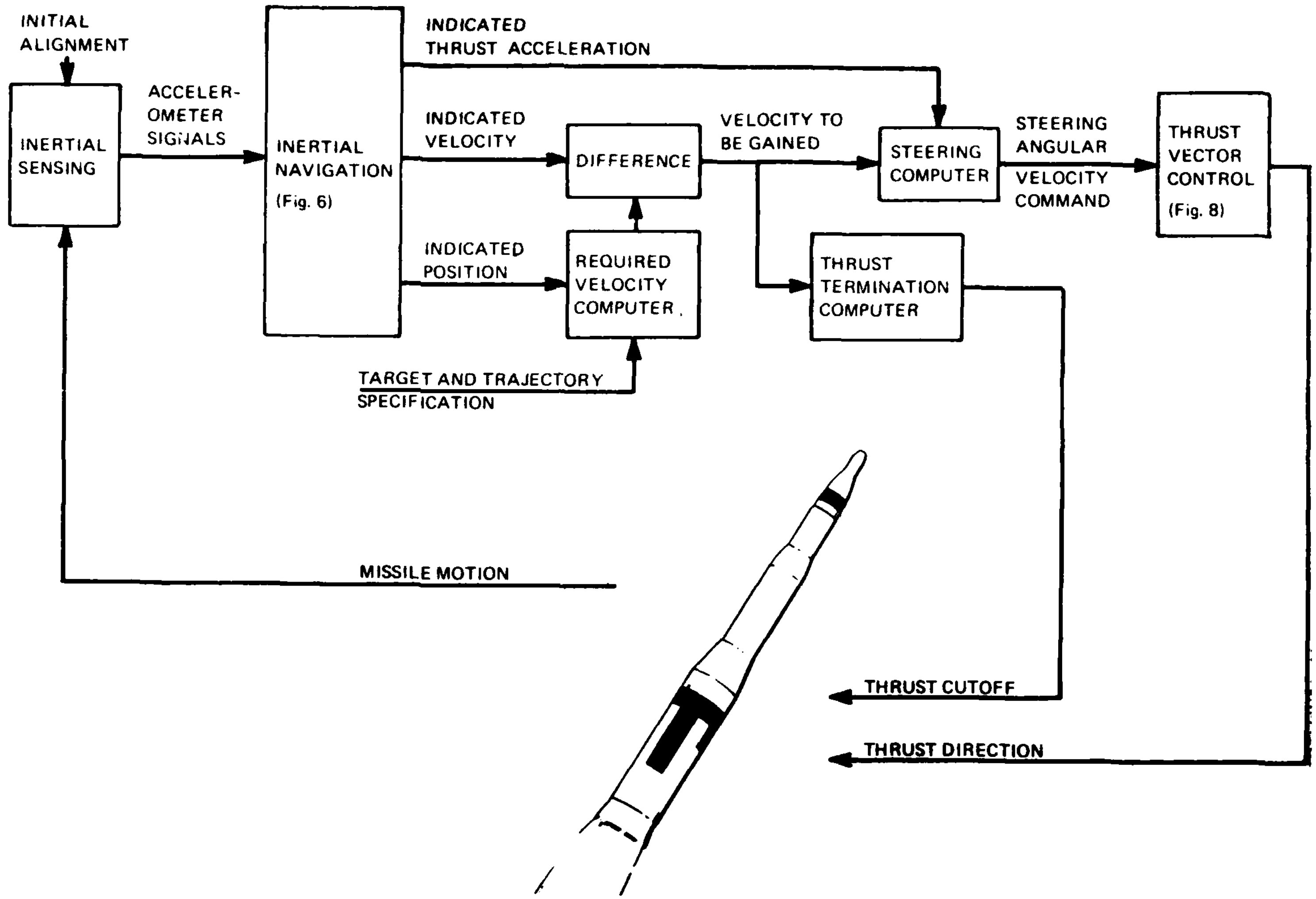


Figure 5-7. Inertial guidance.

direction of the applied rocket acceleration. For liquid-fueled rockets, the engine is relatively small, and the use of an engine swivel or gimbal arrangement to deflect these forces from the center of the mass achieves control torques. For solid-fueled engines, the thrust chamber contains all the unburned propellant, and a gimbal mounting of the whole chamber is not practical. Other means of torque control such as jet vanes, jetavators, nozzle swivel, or gas injection can deflect the hot gas stream without moving the thrust chamber itself.

These methods provide control torques about pitch and yaw axes perpendicular to the long axis, the roll axis, of the rocket. Multiple engines or nozzles at the base of the rocket are common, and engineers can offset the thrust from each engine or nozzle differentially to provide torque control about the roll axis. Otherwise, they can array special small thrusters on the side of the missile for roll control.

Figure 5-8 illustrates a typical thrust vector control loop. The input to this system is the signal from guidance proportional to the desired angular velocity of the thrust direction to reduce the thrust-direction error towards zero. This angular velocity command is compared with the actual missile angular velocity presently existing. Rate gyroscopes measure this actual missile angular velocity. The difference becomes the missile-angular-velocity error that the box labeled dynamic control amplifies and treats with dynamic compensation filtering. The output of this box is the command to the torque control deflecting the thrust such as described previously. The missile responds to the applied torque according to its moments of inertia and the other dynamic properties. The rate gyroscopes sense the resulting angular motion to close the control loop.

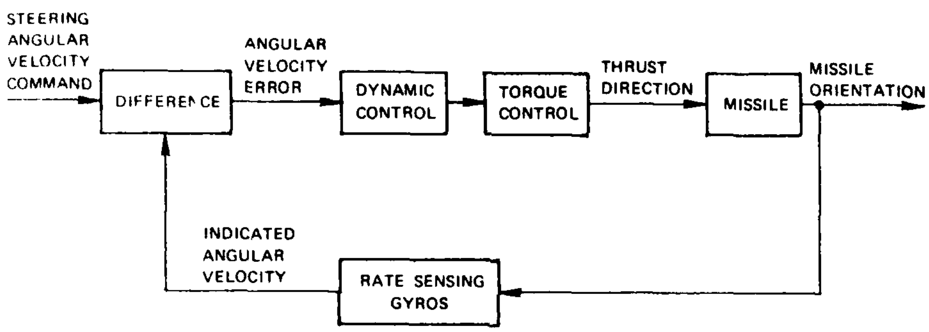


Figure 5-8. Thrust vector control.

The dynamic compensation of such loops can be a complex design process. We can treat the missile as a rigid body in design only rarely, and the effects of body bending and fuel and oxidizer sloshing in tanks can cause large and usually destabilizing torques. Another effect with a very descriptive name is that of engine "tail wagging," where large gimballed engines have moments of inertia comparable to the rest of the missile. Commands to move the engine (the tail) result in considerable motion of the missile itself (the dog). And finally, in the flight through the atmosphere, the effects of aerodynamic forces are considerable and destabilizing.

The thrust-vector-control design must recognize these destabilizing effects so that the dynamic error in following guidance commands is sufficiently small. This requirement is particularly tight near thrust termination. If the missile has high-angular velocities at this time or is not following steering commands accurately, then all three components of the desired velocity will not pass simultaneously through their correct values for engine cutoff. Residual velocity error causes miss at the target if not subsequently corrected.

Recapitulation

A booster rocket propels a ballistic missile to a high velocity above the atmosphere, where the

missile releases the warhead to coast free in a ballistic path before it reenters the atmosphere. Guidance of this weapon to hit a specific target occurs only during the rocket-boost phase, which extends over only the first few minutes of flight. Guidance measurements utilize inertial-sensing techniques. Inertial sensing is the application of Newton's laws as they operate in gyroscopes and accelerometers to measure all components of missile rotational-motion changes and all components of missile translational-motion changes due to all forces on the missile except gravity. Navigating, or determining missile position and velocity by inertial sensing, requires the analytical determination of that component of missile motion due to the force of gravity acting upon it. Moreover, since the inertial sensing can measure only changes in the missile motion, the initial conditions of position (launch location) and velocity must be provided externally before launch. Guidance calculations determine the direction to steer the rocket so that the velocity and position approach those conditions compatible with a free-fall coast to the target identified to the guidance system at launch. The rocket-autopilot thrust-vector-control system responds to the guidance steering commands to control the direction of rocket thrust with respect to the missile in a stable fashion. When the missile achieves the proper conditions of position and velocity, the guidance system signals termination of rocket propulsion and releases the warhead. Now flying under the influence of gravity alone, the warhead coasts in the vacuum of space in an elliptical path until it enters the earth's atmosphere where it may slow considerably before being triggered to explode at the target.

IN-ORBIT ATTITUDE CONTROL

The second item for expanded discussion is in-orbit attitude control. We use the NASA Orbiting Astronomical Observatory (OAO) for the illustration.

Orbiting Astronomical Observatory Attitude Control

Two astronomy experiments are aboard the OAO-C. First is a Princeton University experiment (PEP) that uses a 32-inch telescope to examine interstellar media by measuring the media absorptive characteristics in the ultraviolet spectrum using stars as light sources. The second experiment is the University College, London, experiment that uses three small telescopes to measure X rays. Both experiments require attitude-hold control during a measurement and attitude-change control (slewing) when it is time to observe in a new direction. *Figure 5-9 depicts* the basic operation of both modes of attitude control.

Orbiting Astronomical Observatory Attitude Hold

The basic requirement for attitude hold is maintaining the vehicle at or very near a desired orientation in inertial space, for example, orientation referenced to the "fixed" stars. Measurement of small rotation-angle deviations from the prescribed attitude is necessary for control.

For the PEP experiment, the 32-inch telescope measures two attitude-error angles about the pitch and yaw axes of the vehicle. These angles are the small rotations of the telescope boresight, for example, the vehicle roll axis away from a star line of sight (see fig. 5-10). An inertial-reference unit (IRU) measures error angle about the roll axis. The IRU is a set of three gyroscopes strapped down to the satellite with their three input axes aligned with the yaw, pitch, and roll axes. During periods when the telescope cannot observe the star, the IRU measures pitch and yaw error angles. For the X-ray experiment, the IRU measures error angles for all three axes.

Given the measured error angles, the attitude-hold control system rotates the vehicle to drive the error angles to zero. Fine momentum wheels (FMW) and the magnetic unloading system (MUS) accomplish these attitude correction rotations.

Each FMW can rotate the vehicle about the wheel axis of revolution. The physical principle behind use of a FMW is the conservation of angular momentum. The following equation expresses this principle: $I_v\omega_v + I_w\omega_w$ = constant. In this equation I_v and I_w are the vehicle and wheel moments of inertia about the controlled axis, and ω_v and ω_w are the vehicle and wheel angular velocities about

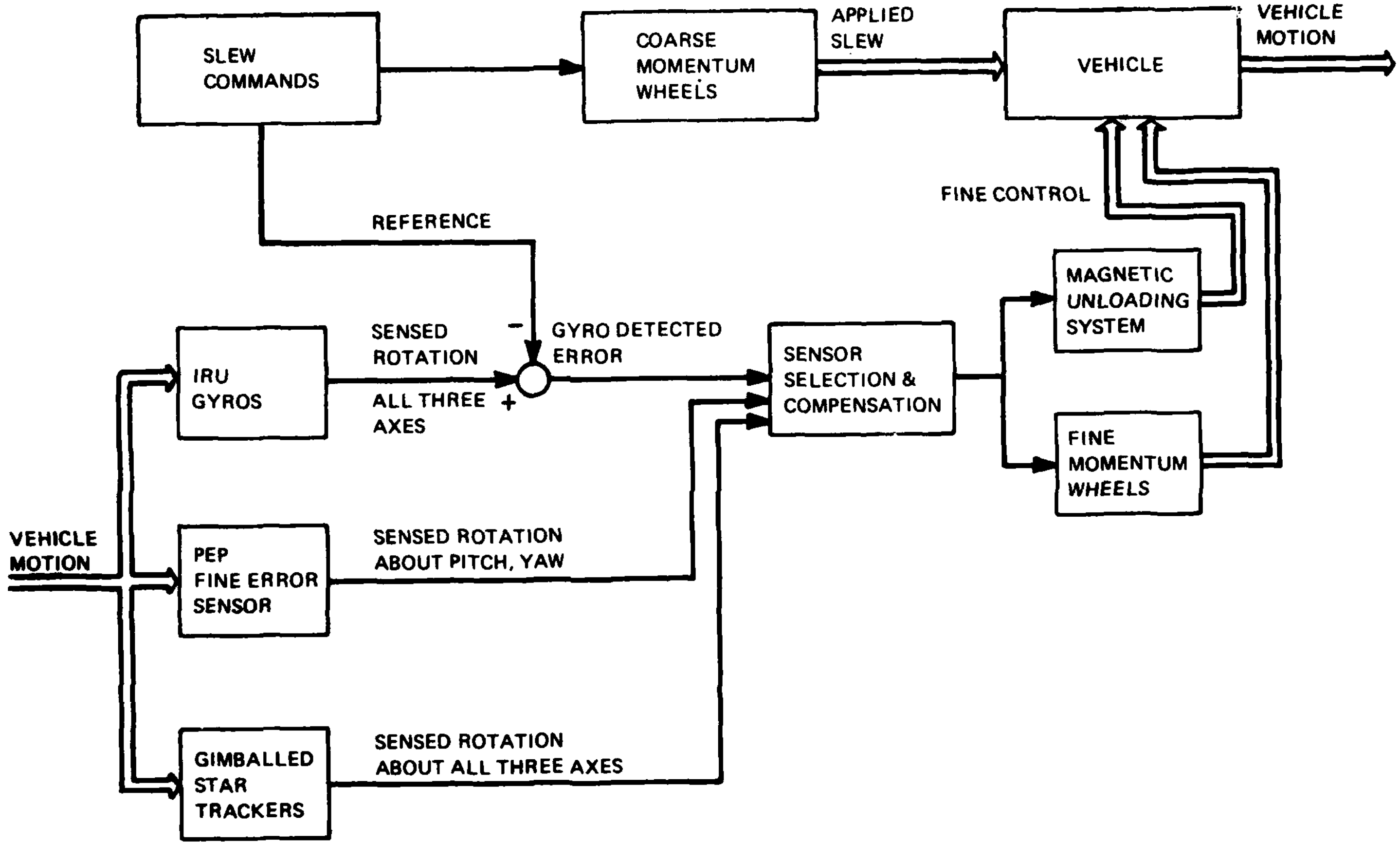


Figure 5-9. OAO attitude control.

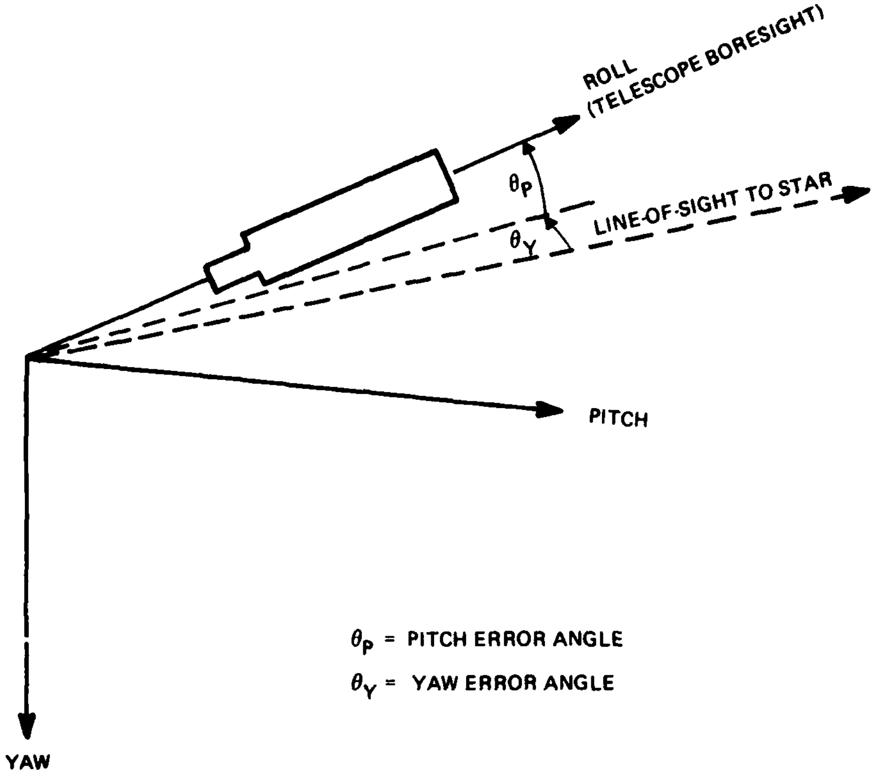


Figure 5-10. Attitude-error angles measured by PEP.

this axis with respect to inertial space. By increasing ω_w, the vehicle angular velocity will change to satisfy the equation and, in so doing, rotate the vehicle as desired. By computing the time integral of this equation (the angles rather than angular velocity), we can rotate the vehicle through the negative of its error angle about the control axis. The primary cause of attitude deviations that FMWs control is torque by gravity gradients.

The primary function of the magnetic unloading system is unloading angular momentum from the FMWs. In the process of controlling vehicle attitude, the FMWs tend to approach their maximum wheel speeds. To prevent these maximum, or saturation, wheel speeds from occurring, the MUS continually applies compensation torque to the vehicle such that the wheel speeds do not saturate. In addition to this desaturation function, the MUS torque provides additional fine attitude control of the vehicle. Passing electric currents through selected wire coils in the vehicle obtains the MUS torques. Current in the coils interacts with the earth's magnetic field and produces a torque on the vehicle.

Slewing is rotating the vehicle from one attitude to another. Given present attitude and desired new attitude, ground-based computers determine the slew commands needed. The slew commands are a set of three rotations that the vehicle takes one at a time. Each rotation is taken about a specified axis.

The slew commands form what we know as a Euler-angle rotation sequence. We can find the definition of Euler angles in most classical mechanics text books. It is interesting to note that we can define 24 unique Euler sequences for a given attitude change. We consider all 24 in the ground computations and eliminate some because the large telescope would swing by too bright an object such as the sun, earth, or moon. Of the remaining possible sequences, we use the one requiring minimum control. Minimum control occurs for the one having the smallest sum of angles of rotation.

We accomplish slewing with coarse momentum wheels (CMW) with fine control from the FMW and MUS. We must express each of three command rotation angles as its equivalent number of revolutions of the appropriate CMW. Then each CMW rotates through its commanded number of revolutions, with possibly one or two error revolutions. The IRU gyros sense these errors, plus any other errors in the slew. The FMWs and MUS compensate for these areas, yielding the final desired slew.

HARDWARE IMPLEMENTATIONS

In previous sections, we described the concept of inertial guidance, navigation, and control. We made reference to inertial sensors and systems, star trackers, and other means of determining and controlling the attitude of a booster or an orbiting satellite. No discussion of GNC is complete without a brief mention of the various pieces of hardware employed to achieve mission requirements.

As a point of reference, consider that the first military application of GNC for space vehicles was the German V-2 used in World War II. Since then, we have made great strides in accuracy and reliability to the point that we can achieve extreme accuracy in orbit insertion or warhead delivery. One of the next major breakthroughs will be the control and pointing of a line of sight from orbit to accuracies equivalent to those associated with a ground-based astronomical observatory. The large space telescope program of NASA has this as a prime design goal. Some hardware implementations that we are using now for GNC cover a technology span from radar/radio trackers to completely self-contained inertial systems requiring no external updates.

Radar/Radio Tracking

During the early days of this nation's space effort, launch control was achieved by radar tracking of the vehicle. Figure 5-11 shows a typical system used for launch control. This type of control was not completely satisfactory. As the state of the art improved, inertial control replaced it. In today's launch operations, we use radar/radio tracking as a monitor to confirm that the spacecraft is achieving the desired trajectory. It provides the range safety officer with an excellent tool from which to make his decision to destruct if necessary. NASA uses radio-energy tracking for all its deep space probes because extreme accuracy is not a problem, and there is sufficient time between midcourse guidance corrections to compute the required course changes.

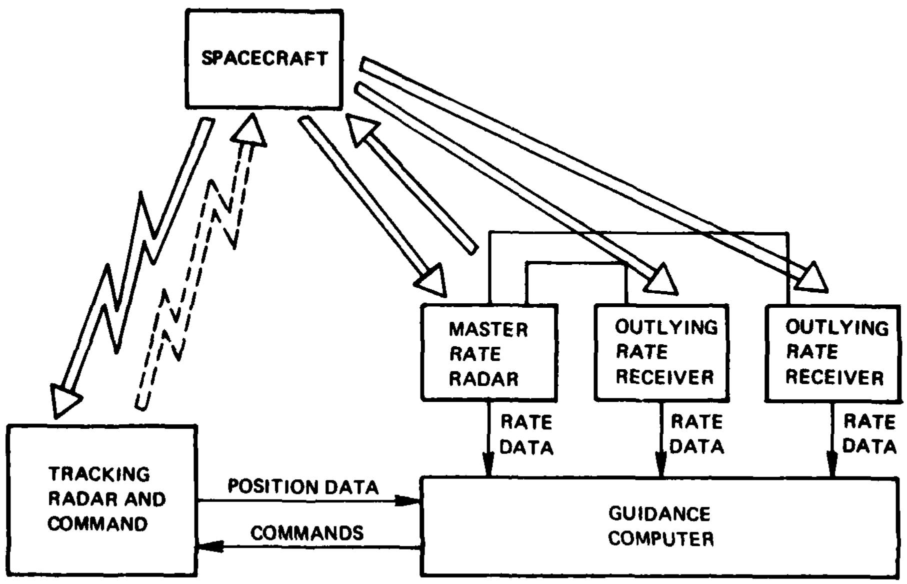


Figure 5-11. Radio command guidance.

Inertial Systems/Sensors

With the exception noted pertaining to deep space probes, all of today's space missions rely on inertial systems for GNC for launch. Depending on the mission, the spacecraft will use inertial sensing for the orbit phase. For reentry, the spacecraft may use inertial sensing depending on the mission and the aerodynamic characteristics of the vehicle. The command module for Apollo maintained the inertial system in operation until splash down. At present there are three mechanizations in use, namely, gimballed, strapdown, and floated ball. Each of these has its own advantages and disadvantages depending on mission requirements: cost, performance, and reliability, to name a few.

The gimballed system is the one used most frequently because it represents the one with the greatest technology background. As we described in previous sections, it consists of three single-degree-of-freedom gyros and a pair of two-degree-of-freedom gyros and three accelerometers mounted on a platform in such a way that the instrument input axes are mutually perpendicular to sense motion about or along the X, Y, and Z axes of the inner gimbal. The gyros and associated servo loops drive the inner, middle, and outer gimbals to maintain the instrument stable regardless of vehicle angular motion. Each gimbal has an electrical-mechanical readout device to measure gimbal position. We use these gimbal angles as inputs, along with acceleration data, to a computer that generates guidance commands, navigation information, and compensation for inertial sensor errors. This last function is necessary because no one has made a perfect instrument as yet. However, we can predict its errors. With a knowledge of gyro and accelerometer error, it is possible to compensate for these errors during operation. A classic example of the use of a gimballed inertial system is its use in Minuteman, transport aircraft, and submarines.

As its name implies, a strapdown system mounts all of the sensors directly to the vehicle structure. The output of each set of sensors, gyros, and accelerometers is body angular rotation about the X, Y, and Z axes and specific force along the respective axis. These data are input to a digital computer that, in addition to the normal GNC function, converts body-referenced information to a suitable nagitation frame. The "gimbal system" is resident in the computer as an attitude transformation. Strapdown systems trade mechanical precision and manufacturing cost for a more complex digital computer. The decision process for the selection of gimballed versus strapdown is a complex one involving many factors. In general, we will choose a gimballed implementation if we require high accuracy and cost is not a driving factor.

We can compare the floated-ball configuration to a combination of both the previous implementations. The floated-ball configuration contains the inertial sensors and their associated electronics within a spherical structure that floats in a temperature-controlled sphere in such a way that the instrument ball is neutrally buoyant. Hydraulic torque motors accomplish the centering of the ball with respect to the sphere. These motors make the ball rotate in response to gyro-generated control signals as the vehicle rotates. An associated digital computer provides GNC computations as in the previous implementations, as well as sensor compensation. The great advantage to this configuration is the high degree of protection provided the inertial sensors when they encounter high "g" loads. Figure 5-12 is an exploded view of such a system.

Many applications do not require a three-axis inertial measurement unit or system. As an example, the Saturn launch vehicle for Apollo employs a triple-redundant three-axis rate gyro package to generate body rate information for the flight control autopilot. The triple redundancy provided the required Apollo reliability by voting the outputs of each gyro. If the spacecraft requires more accurate data than it can secure from a rate gyro, we will use a single-degree-of-freedom rate integrating gyro in a torque feedback mode. In this mode, the spacecraft amplifies the gyro output and uses it to drive its null to zero. The current required to do this is a measure of attitude change, and, as it is generally in digital form, provides a convenient input for a digital guidance/control computer.

For a long time, aircraft have used two-degree-of-freedom instruments for autopilot applications and vertical reference. Until the 1970s, they were used very little for space vehicles because of performance limitations. Since then, their performance has improved greatly and many spacecraft

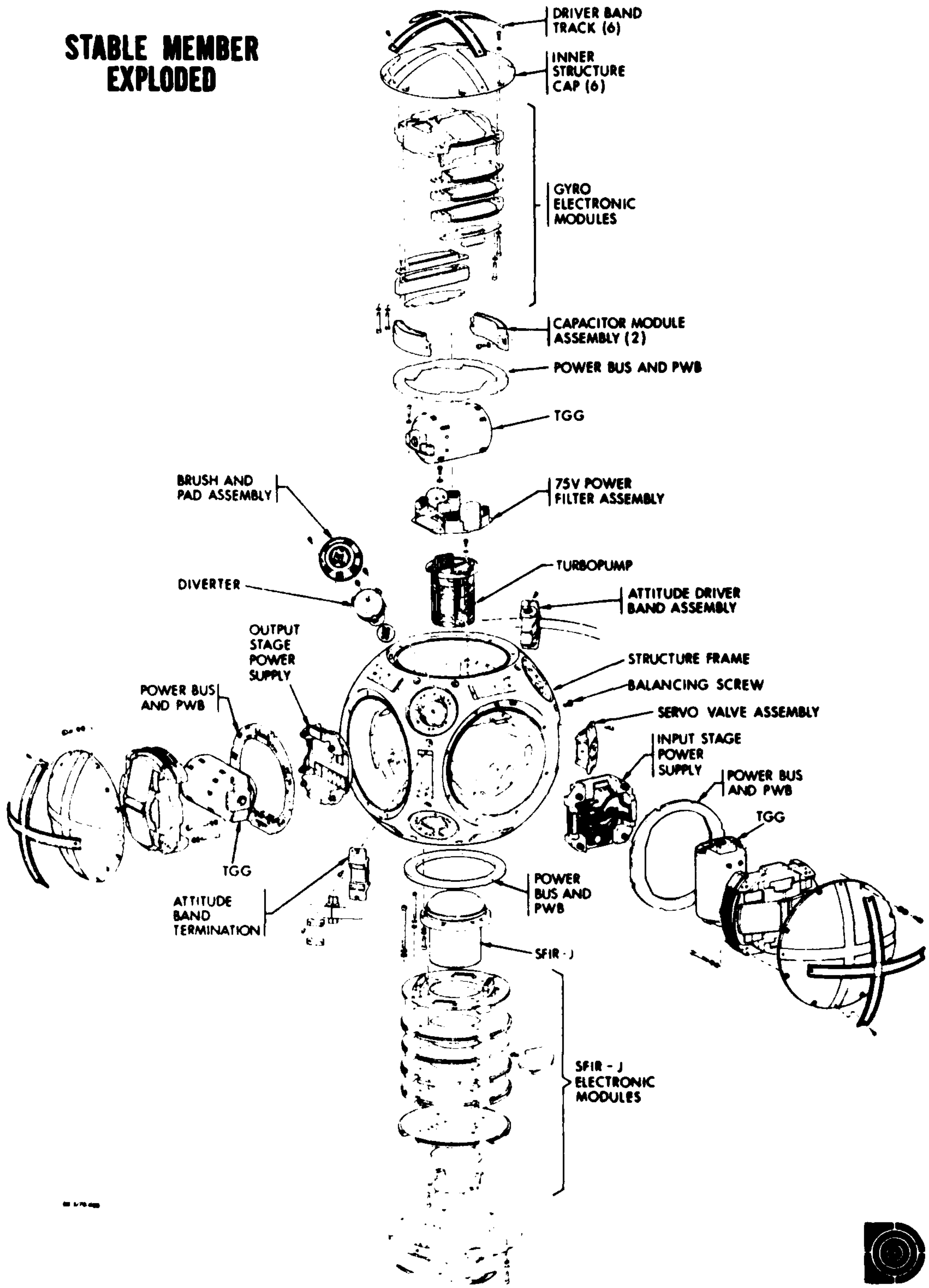


Figure 5-12 Floated-ball configuration

are using them in both gimballed and strapdown configurations. They have the advantage of having one axis with a redundant sensor, which the system designer can use to advantage. If the spacecraft requires high reliability about one axis, engineers can configure the system in such a way that two gyro outputs indicate rotations about the critical axis.

The laser gyro is another sensor receiving increasing notice. This instrument has the potential of equaling the performance of "iron" gyros with the advantage, in theory, of no moving mechanical parts with the exception of a possible dithering subassembly. Aircraft and tactical weapons have flown prototype instruments with encouraging success. Effort is underway to apply a Laser IMU for space application.

Accelerometers for space applications do not differ greatly from those required for aircraft use. Engineers are concentrating their development efforts on improving the accuracy and resolution and on reducing the magnitude of error terms. In general, the sensing element is a damped spring-restrained mass that is driven to null by electrical means. Its output is digital and, depending on the instrument design, will indicate either velocity or acceleration. If the output is velocity, we will call the instrument a velocity meter or an integrating accelerometer. The more accurate instruments take this form.

Finally, industry and DOD laboratories are applying considerable effort to develop multisensor instruments. These are units that provide both acceleration and angular-rate information from one instrument. As of the middle 1970s, engineers had laboratory tested a number of concepts, and fabricated one single-axis platform incorporating a multisensor. The prime motivation behind the development of these instruments was an attempt to reduce overall systems cost. Their performance precludes their application to medium- or high-performance mission requirements.

Star Trackers

These instruments are used to track a star and measure its direction with respect to the base on which the inertial system is mounted. In conjunction with a computer, their purpose is to determine the spacecraft's position and attitude in space. By using a series of measurements, the observers try to compensate for the buildup of errors in an inertial system. As there is no defined horizon, a space sextant (star tracker) measures the angle between two stars and a near body to determine two lines of position. A series of measurements is made using many combinations to secure an accurate position. The sensor commonly used is a vidicon tube whose photoconductive sensor is sensitive to star emissions. In operation, the vidicon is part of a closed-loop servo that automatically tracks and measures the position of a selected star with respect to the inertial system. A series of measurements provides the data from which vehicle position is determined. The process of determining position is similar to that used for aircraft navigation using a bubble sextant and sight reduction tables. In a spacecraft, the process is automatic, and the sight reduction tables convert to stored ephemeris data for the stars of interest.

Horizon Sensors

These instruments are used to sense the earth's horizon or limb, and their output is employed to maintain a satellite's position, generally the pitch axis, fixed with respect to the earth's center. Reaction wheels, control-moment gyros, or gas jets accomplish control of pitch-axis position, depending on the design of the vehicle. The sensor, mounted at the focal plane of an optical system, is sensitive to infrared radiation present at the earth's limb. A rotating shutter or light chopper within the optical system modulates the radiation received by the infrared sensor and serves to generate a phase-sensitive error signal used to control the satellite pitch axis. Generally, this type of sensor is used when stabilization requires moderate accuracy.

SATELLITE TELEMETRY, TRACKING, AND COMMANDING

The three functions involved in satellite control are telemetry, tracking, and commanding (TTC). All satellites regardless of their mission or capability require TTC. Taken in logical order, first, we must be able to locate a satellite in space and track it; second, we must be able to communicate or command the satellite to perform various functions; and third, we must be able to obtain data from the satellite (telemetry).

The heart of satellite telemetry, tracking, and commanding for the DOD is the Satellite Test Center (STC) located in Sunnyvale, California. The STC performs TTC functions through seven remote tracking stations located around the world. Air Force Satellite Control Facility (AFSCF) is the name for the entire network of remote tracking stations (RTS) and the STC.

Telemetry

Satellite telemetry is measurements taken by remote sensors on a satellite and transmitted to a ground station. The two basic classifications of telemetry are state-of-health data and payload data. State-of-health data pertains to the satellite itself. Ground control uses it to determine the operational status of the satellite and its equipment. Payload data pertains to the mission of the satellite.

Tracking

Satellite tracking involves locating a specific satellite in time and space and following its movement as a function of time. The purpose of tracking is to enable the sending of commands, to acquire telemetry, and to provide data for orbit determination. Elevation, azimuth, range, and range rate are the measurements used in satellite tracking to determine the position of a satellite as a function of time.

Commanding

Commanding is the method of controlling the satellite from the ground while the satellite is in the line of sight of a ground station. The STC generates most commands and relays them over land lines, submarine cables, microwave relay, and satellite links to the RTS. Later, upon direction, the RTS sends them to the satellite. Table 5-1 gives a typical list of satellite command functions. The two types of commands are real-time and stored program. The satellite receives and acts on the real-time commands immediately. Stored program commands activate satellite systems and sensors while the satellite is not in line of sight of an RTS. The four primary modes of transmitting these commands are single, block, repetitive, and timed repetitive.

A single command is the transmission of only one command word. An example would be a turn-on command. A block command is a group of single commands represented by a single command number. The repetitive command is a single or block command that the RTS retransmits continuously until there is command verification or the transmission efforts reach a selected number. A timed repetitive command is the least frequently used method of command transmission. Using this method, the RTS transmits a block or single command for a specific period of time.

Table 5-1.
Satellite Command Functions

On/off	*Other*
Payload or sensor	Payload calibration
Main engine	Eject horizon fairings
Secondary propulsion system	Telemetry from ascent to orbit mode
PCM telemeter unit	Tape recorder reload
Downlink transmitter modulation	Tape recorder reproduce
Select decoder no. 1	Tape recorder high-speed mode
Select decoder no. 2	Tape recorder advance
Backup decoder	Disable tape recorder
Solar array panels	Paired solar array panels
(entirely or by sections)	Fire gyro squid (variety type)
	Telemetry data rate selector
	Reentry sequence initiation
	Orbit adjust/thruster control
	Select on-orbit antenna

BIBLIOGRAPHY

Bate, R. R., D.D. Miller, and J.E. White. *Fundamentals of Astrodynamics.* New York: Dover, 1971.

Battin, R. H. *Astronautical Guidance.* New York: McGraw-Hill, 1964.

Britting, K. *Inertial Navigation Systems Analysis.* New York: Wiley, 1971.

Broxmeyer, C. *Inertial Navigation Systems.* New York: McGraw-Hill, 1964.

Ford Aerospace and Communications Corporation. *SCF Telemetry, Tracking, and Commanding System Orientation.* SCF Course 116. Sunnyvale, Calif.: Ford Aerospace and Communications Corporation, January 1983.

Kaplan, M. H. *Modern Spacecraft Dynamics and Control.* New York: Wiley, 1976.

Leondes, C. T., ed. *Guidance and Control of Aerospace Vehicles.* New York: McGraw-Hill, 1963.

Miller, J. E., ed. *Space Navigation Guidance and Control.* New York: Circa Publications, 1966.

Pitman, G. R., ed. *Inertial Guidance.* New York: Wiley, 1962.

Thomason, W. T. *Introduction to Space Dynamics.* New York: Wiley, 1961.

Wrigley, W. Hollister and W. G. Denhard. *Gyroscopic Theory, Design and Instrumentation.* Cambridge, Mass.: MIT Press, 1969.

Chapter 6

ACTIVE US SPACE LAUNCH VEHICLES AND UPPER STAGES

The satellites or payload going into orbit are mounted on top of a launch vehicle and an upper stage. We discussed orbital mechanics in chapter 2. The launch vehicle provides the large amount of energy required to escape the earth's gravity and places the payload into an elliptical orbit. The apogee of the ellipse depends on the required altitude of a given payload and varies from mission to mission. The upper stage, or stages, provides the energy to place the payload into its final orbit. If the final orbit will be circular in nature, the designers can use a single upper stage. Energizing takes place when the payload reaches the apogee of the ellipse. The result is a final orbit in a circle with a radius equal to the apogee of the original ellipse. If the final orbit will be elliptical in nature, it will require a second upper stage or second energizing of a single upper stage. The second upper stage provides the energy to place the payload into an elliptical orbit that has a perigee equal to the radius of the circular orbit previously described. It has an apogee determined by the amount of energy that the second upper stage provides. This chapter describes the launch vehicles and upper stages used by the United States to orbit payloads. The bibliography at the end of this chapter provides further information on documents dealing with the capabilities of particular systems.

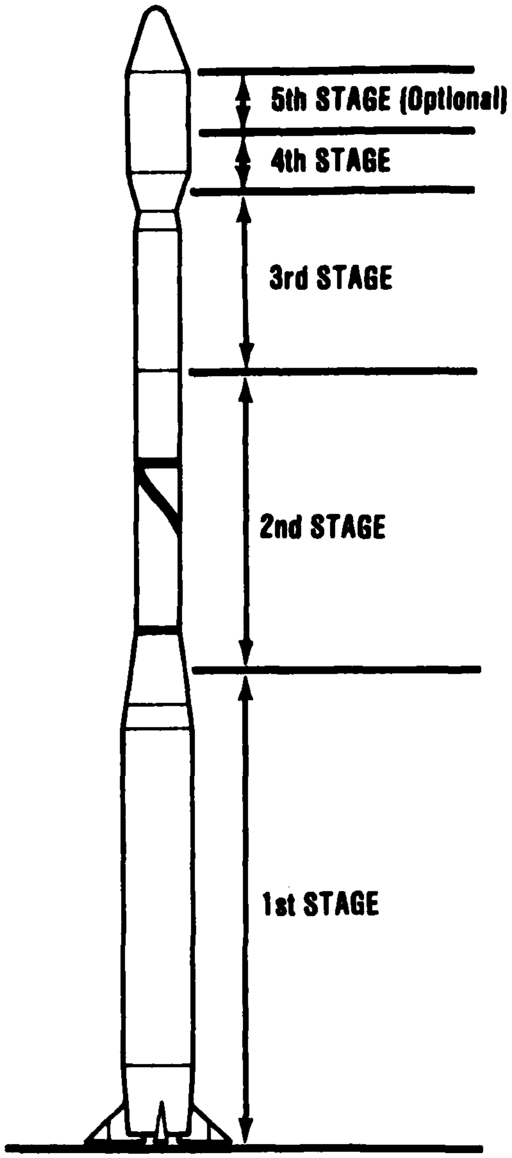
SPACE LAUNCH VEHICLES

The Scout, Delta, Atlas, and Titan are current US space launch vehicles. The Air Force and the National Aeronautics and Space Administration (NASA) plan to use these vehicles throughout the 1980s and then phase them out as the Space Transportation System becomes fully operational from both the Eastern Space and Missile Center in Florida and the Western Space and Missile Center in California. The Ascent Agenda, Centaur, Delta Second Stage, and Transtage are upper stages used with the launch vehicles to provide the necessary thrust for placing payloads into their final orbits. The Payload Assist Module and the Inertial Upper Stage are upper stages used with both the Space Transportation System and space launch vehicles.

Scout

The Scout is the smallest of the active launch vehicles and can place a payload of 181 kilograms (300 pounds) into a circular orbit of 556 kilometers (300 nautical miles).* Used since the early 1960s, it is commonly a four-stage vehicle that weights 18,144 kilograms (40,000 pounds) and generates 445 kilonewtons (100,000 pounds) of thrust at liftoff. A fifth stage was developed to place small payloads into a highly elliptical orbit. The Scout shown in figure 6-1 consists of the Algol IIIA first stage,

*All orbit and payload figures in this chapter are approximations.



STAGE IV

Height	m (feet)	1.2 (4.0)
Diameter	m (feet)	.7 (2.1)
Thrust	kn (pounds x 1,000)	25.6 (5.8)

STAGE III

Height	m (feet)	3.4 (11.2)
Diameter	m (feet)	.8 (2.5)
Thrust	kn (pounds x 1,000)	83.1 (18.7)

STAGE II

Height	m (feet)	6.3 (20.7)
Diameter	m (feet)	.8 (2.6)
Thrust	kn (pounds x 1,000)	285.2 (64.1)

STAGE I

Height	m (feet)	9.4 (30.8)
Diameter	m (feet)	1.1 (3.8)
Thrust	kn (pounds x 1,000)	431.1 (96.9)

Figure 6-1. Scout.

Castor IIA second stage, Antares IIIA third stage, Altair IIIA fourth stage, and an Apogee Kick Motor fifth stage. The Apogee Kick Motor gets its name from the face that it becomes energized when the payload is at apogee, thus providing thrust or "kick" to place the payload into its final orbit. All stages of the Scout used solid propellant to generate thrust. A high-temperature device ignites the solid propellant and, as it burns, the propellant produces a gas that expels through a nozzle at the rear of the vehicle. The gas expands as it is expelled and provides the thrust required for propulsion. To ensure application of thrust in the proper direction, a guidance and control system provides attitude reference and control signals necessary for vehicle stability.

The guidance and control system (see chapter 5) uses an inertial guidance system for the first three stages and a spin-stabilized system for the fourth and fifth stages. A launch controller in a reinforced

structure called the blockhouse transmits a direct electrical signal that ignites the first stage. Commands given by a guidance timer on the launch vehicle ignite the second through the fifth stages.

In addition to the guidance and control system, the Scout's instrumentation system includes equipment for monitoring the subsystems during prelaunch checkout and countdown, telemetry monitoring of critical performance parameters, and receiving and decoding of destruct information. A destruct system provides a means of destroying the vehicle if any malfunction presents a safety hazard. A high-energy charge destroys the first three stages before the vehicle reaches the ground.

Delta

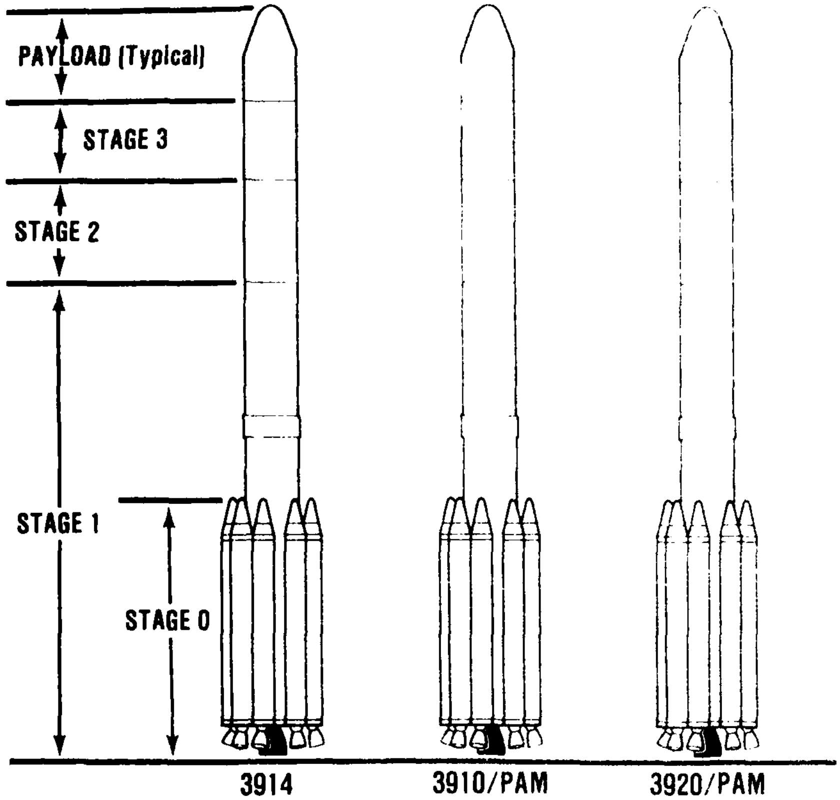
The original Delta created by NASA as an intermediate size launch vehicle in 1959 was capable of placing a payload of 272 kilograms (600 pounds) into an orbit of 185 kilometers (100 nautical miles). Later models of the Delta are capable of placing a payload of 1,270 kilograms (2,800 pounds) into an orbit of 296 by 35,786 kilometers (160 × 19,323 nautical miles) See figure 6-2.

The original Delta consisted of three stages; namely, a Thor first stage and two upper stages from the Vanguard program. US engineers derived the first stage from the Thor, which was developed in 1958 and deployed as an intermediate range ballistic missile in England. The United States first used the basic Thor as a space launch vehicle in 1958 (Pioneer I). The basic Thor was used for the final time in 1980 (Block 5D-1 of the Defense Meteorological Satellite Program). The basic vehicle was modified to increase its payload capability for use in the Delta configuration. They modified the Vanguard second stage, which used liquid propellants, to improve its guidance system and its control during unpowered portions of the flight. They used the third stage motor without modification.

From this baseline configuration, the Delta vehicle progressed through a remarkable series of modifications to increase its payload capabilities. Modifications included addition of rocket motors with increased thrust, new engine technology, and an upgraded guidance system. Usually, the engineers employed modifications to the Delta only when components became available from other programs. They would get components by adding to orders from other programs, or by making modifications. This approach ensured minimum costs of modifying the Delta to meet new mission requirements. The first two models, A and B, followed this philosophy by increasing the first stage thrust and lengthening the second stage propellant tank, both minor modifications. The C and D models were the first to incorporate thrust augmentation during first stage operation. Thrust augmentation included bolting three Castor I solid-propellant motors to the Thor first stage. The motors were ignited at lift-off and jettisoned after they consumed their propellant. These versions incorporated an improved third stage solid-propellant motor from the Scout program.

The Delta E began using a restart capability in the second stage. Combined with larger second stage propellant tanks, this capability allowed the Delta to place payloads into higher circular and elliptical orbits. The spacecraft could use the single second stage like two upper stages since the restart capability could provide additional thrust at two different points in the orbit. Performance increased again when a motor that carried more propellant replaced the third stage solid motor. The next major change came with the Delta L.

Development of the Long Tank Thor in 1968 increased the tank sizes in the first and second stages by eliminating a tapering on the upper portion of the Thor. This modification permitted expansion of the vehicles to a diameter of eight feet and increased tankage in both stages. The L, M, and N models used the Long Tank Thor, replaced the thrust augmentation of the Castor I with the Castor II solid-propellant motor, and increased the number of motors from three to six. The Delta M6 had six Castor II motors and was capable of placing a payload weighing 454 kilograms (1,000 pounds) into an orbit 296 × 35,786 kilometers (160 × 19,323 nautical miles). These increases were required to meet the need for launching communication satellites into the orbit known as the geosynchronous or geostationary orbit. The geostationary orbit is an orbit that allows a satellite to remain over the same location on the earth surface at all times. This orbit has its greatest use for communication and relay satellites used to acquire data between overseas locations and the continental United States. The Delta saw another series of major changes in 1972. The Thor



		3914	3910/PAM	3920/PAM
STAGE III				
Height	m (feet)	2.1 (7.0)	2.0 (6.7)	2.0 (6.7)
Diameter	m (feet)	.9 (3.1)	1.3 (4.2)	1.3 (4.2)
Thrust	kn (pounds x 1,000)	65.8 (14.8)	84.5 (19.0)	84.5 (19.0)
STAGE II				
Height	m (feet)	5.0 (16.4)	5.0 (16.4)	5.0 (16.4)
Diameter	m (feet)	2.4 (8.0)	2.4 (8.0)	2.4 (8.0)
Thrust	kn (pounds x 1,000)	41.8 (9.4)	41.8 (9.4)	41.8 (9.4)
STAGE I				
Height	m (feet)	22.5 (73.7)	22.5 (73.7)	22.5 (73.7)
Diameter	m (feet)	2.4 (8.0)	2.4 (8.0)	2.4 (8.0)
Thrust	kn (pounds x 1,000)	921.0 (207.0)	921.0 (207.0)	921.0 (207.0)
STAGE 0				
Height	m (feet)	11.7 (38.3)	11.7 (38.3)	11.7 (38.3)
Diameter	m (feet)	.9 (3.0)	9 (3.0)	.9 (3.0)
Thrust	kn (pounds x 1,000)	378.0 (85.0) ea	378.0 (85.0) ea	378.0 (85.0) ea

Figure 6-2 Delta

was strengthened structurally to accommodate nine Castor II motors. Engineers added the Transtage engine developed for the Titan launch vehicle to the second stage and replaced the numerical designations with letter designations. Space program personnel knew these configurations as the Delta 903 and 904. Late in 1972, the four-digit number replaced the numerical designation

The first of these designations was the Delta 1900 series, which had a larger third stage motor for increased apogee kick. The Delta 2900 series replaced the MB-3 engines of the first stage with more powerful RS-27 engines, and the Delta 3900 series replaced the Castor II motors with larger and more powerful Castor IV motors. A typical 3900 series Delta weighs 193,233 kilograms (426,000 pounds) at lift-off and develops 2,807 kilonewtons (631,100 pounds) of thrust. Figures 6-2 and 6-3 show current configurations of the 3900 series Delta and identifies the numbers in the four-digit code.

Four-Digit Designator

1st Digit — Type of Augmentation/First Stage
2nd Digit — Quantity of Augmentation Motors
3rd Digit — Type of Second Stage
4th Digit — Type of NASA Third Stage

Delta 3914

NOTE: **When PAM is used, fourth digit is zero and PAM is added to designator — for example, 3910/PAM.**

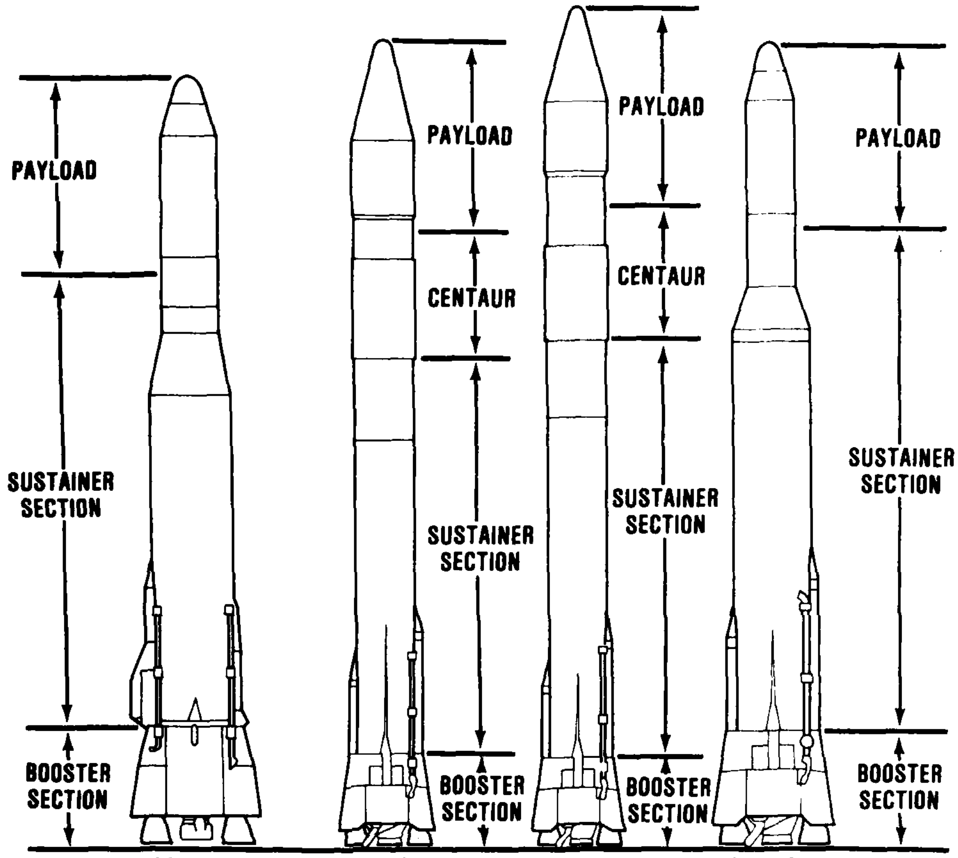
Numbers Used as Designators

- First Digit:
  - **2 — Castor II Augmentation, Extended Long Tank**
  - **3 — Castor IV Augmentation, Extended Long Tank**
- Second Digit:
  - **3 — Three Motors**
  - **9 — Nine Motors**
- Third Digit:
  - **1 — Standard Second Stage (10,000-pound propellant, TRW LEM-D engine)**
  - **2 — Uprated Second Stage (13,200-pound propellant, AJC ITIP engine)**
- Fourth Digit:
  - **0 — No NASA Third Stage**
  - **3 — TE-364-3 Third Stage (1,440-pound propellant)**
  - **4 — TE-364-4 Third Stage (2,300-pound propellant)**

Figure 6-3. Delta nomenclature.

Atlas

Development of the Atlas began in 1946, but the government cancelled the program in 1947 and resurrected it in 1951 (see fig. 6-4). The first test and flight of the Atlas took place in December 1957. The Atlas B had 10 successful tests between 1958 and 1959, and it served as the launch vehicle for Project Score, which relayed President Eisenhower's Christmas message in 1958. The Atlas C used an improved guidance system and carried an operational reentry vehicle in tests conducted during 1958 and 1959. The Atlas D, the prototype for the operational system used a ground-based guidance system but carried the all-inertial guidance system that the Atlas E would use. The missile satisfied all research and development goals and became operational by August 1959. In September 1959, a crew from the Strategic Air Command marked the initial operational capability by launching an Atlas D from the Pacific Missile Range. NASA used the Atlas D to launch the first test flight of Project Mercury during the same month. All of the Mercury flights between 1962 and 1963 used modified



	ATLAS-E	ATLAS-D	ATLAS-G	ATLAS-H
Height m (feet)	20.6 (67.3)	20.2 (65.9)	22.2 (72.6)	20.2 (65.9)
Diameter m (feet)	3.0 (10.0)	3.0 (10.0)	3.0 (10.0)	3.0 (10.0)
Thrust kn (pounds x 1,000)				
Liftoff	1,748.1 (393.0)	1,645.8 (370.0)	1,681.3 (378.0)	1,681.3 (378.0)
Sustainer	253.5 (57.0)	266.8 (60.0)	266.8 (60.0)	266.8 (60.0)

Figure 6-4. Atlas.

Atlas D vehicles for the first US manned orbital flights. Development of the Atlas E and F continued during the same period. The test firing of the Atlas E took place in October 1960. It met all test objectives by May 1961. Testing of the Atlas F began in August 1961 with completion coming by the end of 1962. This marked the end of the five-year test program.

The military deployed operationally Atlas D, E, and F as intercontinental ballistic missiles between 1960 and 1966. They phased out the Altas D in 1964. Shortly thereafter, the Minuteman missile replaced the E and F models.

The spacecraft designers changed the design of the basic Atlas very little over the years. The Atlas is a liquid-propellant vehicle that includes a booster section and a sustainer section. The booster section consists of two high-thrust engines. They ignite at lift-off and the craft jettisons them

approximately two minutes into the flight. The sustainer section has a single engine that ignites at lift-off and operates throughout the flight. The Atlas E uses MA-3 engines on the booster and the sustainer. Other configurations use MA-5 engines. However, the booster engines have been modified to provide the higher level of thrust needed for the lift-off phase of the mission.

Current configurations of the Atlas used for space missions are the Atlas SLV-3D, Atlas G, and Atlas H. Engineers have overhauled completely the original Atlas E vehicles, which stood deployed on alert in the 1960s. The space program is using them as space launch vehicles today. As an example of Atlas performance, the Atlas H weighs 132,723 kilograms (292,600 pounds) at lift-off and generates 1,948 kilonewtons (378,000) of thrust.

Titan

Titan, the largest launch vehicle in the US inventory, is capable of launching a payload weighing 27,600 pounds (12,520 kilograms) into an orbit of 185 kilometers (100 nautical miles). Like the Delta, the Titan has a long history of modification and change that led to its current configuration (see fig. 6-5).

The Titan I was a two-stage liquid-propelled intercontinental ballistic missile that saw its first launch in 1959 and its last launch in 1965. The Titan II, first launched in 1962, replaced it. The military deployed the Titan II later the same year. Although it remains in the nation's inventory of strategic systems, current plans are to deactivate it by 1987. The Titan II began its life as a space launch vehicle in 1965. The Gemini program used it for 10 highly successful launches between March 1965 and November 1966. Engineers modified it only to ensure the safety of the Gemini crew.

Development of a third generation Titan began in 1961 when the need for a larger payload capability became evident. The Titan III was a two-stage liquid-propellant vehicle that employed two solid-propellant motors to augment the thrust capability of the basic vehicle during lift-off. When a vehicle is launched without the solid-propellant motor, using only the two liquid stages, it is known as a core-only vehicle. They last launched the Titan IIIB core-only vehicle from the Western Space and Missile Center in the mid-1970s. The Titan IIIC used the solid-propellant motors. The Eastern Space and Missile Center was the site of its last launching in 1982. The Titan IIID, launched from the Western Space and Missile Center, was similar to the Titan IIIC. Since it did not use an upper stage, the engineers transferred its avionics to stage I and II.

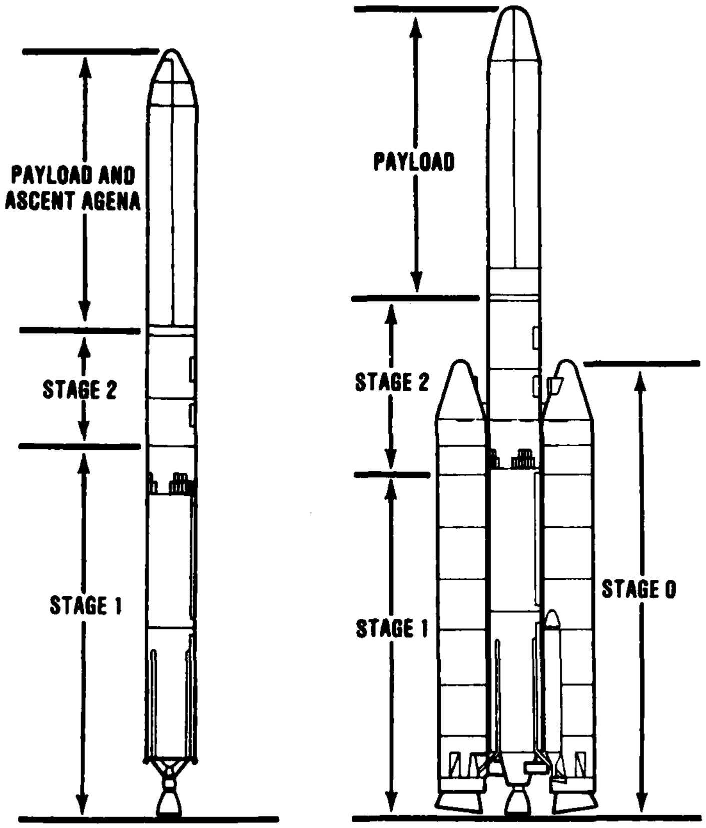
The Titan IIIE was similar to the Titan IIID with the biggest difference being the inertial guidance system's replacement of the radio guidance. It was packaged in the upper stage. This stage was the Centaur D-1T, which provided the thrust capability required to accomplish interplanetary missions. These missions included the Viking missions to Mars, the Voyaer missions to Jupiter and Saturn, and the Helios solar observation missions. The last launch of the Titan IIIE took place at the Eastern Space and Missile Center in 1977.

Another Titan III vehicle was the Titan IIIM designed to launch the Manned Orbiting Laboratory (MOL) from the Western Space and Missile Center. Although NASA cancelled the MOL program, the design for the IIIM was the forerunner of the fourth generation Titan, the Titan 34 series.

The Titan 34B and D are the two versions of the Titan currently used as launch vehicles. They have an improved guidance system, increased structural capability to support heavier payloads, and a larger payload fairing system that allows more space for larger payloads. We launch the Titan 34B from the Western Space and Missile Center and the Titan 34D, user of the inertial upper stage in its first launch in 1982, from the Eastern Space and Missile Center. Since the Titan 34D uses larger solid-propellant motors than the Titan IIID, it has a payload capability that makes it the largest launch vehicle in both size and capability. It weighs 689,300 kilograms (1,519,600 pounds) at lift-off and generates 12,998 kilonewtons (2,920,000 pounds) of thrust.

UPPER STAGES

Upper stage vehicles currently in use by the United States to place payloads into their final orbits are the Ascent Agena, Centaur, Delta Second Stage, Transtage, Payload Assist Module, and Inertial



		TITAN 34B	TITAN 34D/ No Upper Stage
STAGE II			
Height	m (feet)	9.6 (31.3)	9.6 (31.3)
Diameter	m (feet)	3.1 (10.1)	3.1 (10.0)
Thrust	kn (pounds x 1,000)	449.2 (101.0)	449.2 (101.0)
STAGE I			
Height	m (feet)	23.8 (77.8)	23.8 (77.8)
Diameter	m (feet)	3.1 (10.0)	3.1 (10.0)
Thrust	kn (pounds x 1,000)	2,059.4 (463.0)	2,353.0 (529.0)
STAGE 0		N/A	
Height	m (feet)		27.6 (90.4)
Diameter	m (feet)		3.1 (10.2) ea
Thrust	kn (pounds x 1,000)		6,227.2 (1,400.0) ea

Figure 6-5. Titan.

Upper Stage. They use either solid or liquid propellants and normally contain electrical, propulsion, command and control, and guidance systems in a structure that bolts to the payload on the front end and the launch vehicle in the rear. The typical electrical system consists of batteries that provide power for the vehicle and circuits that connect the functional components of the other systems.

The propulsion system uses either solid or liquid propellants. A solid-rocket motor contains the

solid propellant. Therefore, the solid propellant does not require a tank. Once ignited, the propellant burns until all of it is consumed. The solid-propellant system is a one-burn system since it cannot be used for additional firings. However, scientists can design liquid-propellant systems to incorporate a restart feature. These liquid-propellant systems can be monopropellant or bipropellant (see chapter 3). They carry tanks containing the propellants necessary for the mission. A command and control system issues the command that ignites the propulsion system.

The command and control system contains components that receive commands from the ground control station or an on-board guidance computer and relay them to other upper stage systems for execution. The command and control system energizes the other systems by issuing commands through electrical circuits to the propulsion system. These commands control the ignition of the propellant and termination of thrust for liquid systems. It provides commands to maintain the stability of the upper stage throughout the mission.

The guidance system, the "brains" of the upper stage, provides position, velocity, and acceleration data to the command and control system to ensure the issuance of commands at the proper time. It provides data that ground control personnel use to monitor flight performance.

Ascent Agena

The Titan 34B, currently launched from the Western Space and Missile Center, uses the Ascent Agena that was the upper stage for the Titan IIIB from 1966 to 1975. The vehicle is capable of providing the thrust required to place a payload weighing 3,600 kilograms (7,940 pounds) into a circular orbit of 185 kilometers (100 nautical miles). See figure 6-6.

The Ascent Agena consists of propulsion, electrical, guidance and control, and communication systems. The propulsion system includes two tanks that carry the liquid propellant required to develop its thrust and an adapter section that houses the liquid-rocket engine. The system can be operated twice during a single mission. It can provide 75.6 kilonewtons (17,000 pounds) of thrust over a total operating time of 240 seconds. The electrical system supplies and distributes power to operate the vehicle from launch to completion of the mission. The guidance and control system consists of an inertial guidance section that feeds information on position and velocity to the control section that steers both the Titan and the Ascent Agena. The communications system transmits performance data, provides tracking signals, and receives command signals issued from the ground to direct the vehicle.

Centaur

Development of the Centaur began in 1958. Its first launch on an Atlas occurred in 1963. In 1969 and 1972, the Centaur was modified to increase reliability and redundancy (the use of two components to perform the same function and increase the probability of success). See figure 6-7.

The current operational Centaur, the D-1A, consists of an equipment module, propulsion system, and an adapter. The equipment module provides the interface with the payload on the forward end, mounting for the avionics, and an interface with the adapter at the rear end. The propulsion system uses two liquid-propellant rocket engines and has a multiple burn capability. It can produce 133.4 kilonewtons (30,000 pounds) of thrust over a total operating time of 444 seconds. The system has tankage for 13,950 kilograms (30,750 pounds) of propellant. Also housed in the equipment module are the tanks and engines, the inertial guidance system, and the control system that provides for command and control of the Centaur and Atlas throughout the mission. At the rear of the equipment module is the mounting surface that interfaces with the adapter that provides the structural connection between the Centaur and the forward end of the Atlas.

When the Centaur and the Atlas SLV-3D launch vehicles are used together, the Centaur can produce the upper stage thrust required to place a payload of 1,995 kilograms (4,400 pounds) into an orbit of 185 × 35,786 kilometers (100 × 19,323 nautical miles). Although the Centaur is compatible with the Titan IIIE, the Atlas G is the only other launch vehicle that currently uses the Centaur.

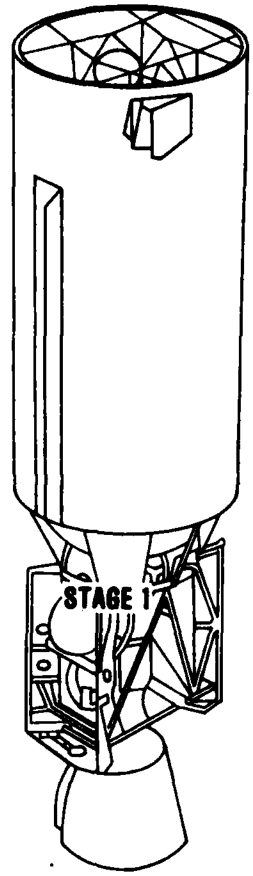


Figure 6-6. Ascent Agena.

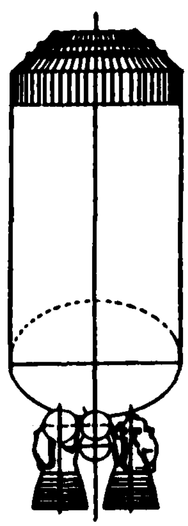


Figure 6-7. Centaur.

Delta Second Stage

The Delta Second Stage comes in two configurations and both are compatible with the Delta 3900 series of launch vehicles. The standard second stage can provide the thrust required to place a payload of 1,111 kilograms (2,450 pounds) into an orbit 296 × 35,786 kilometers (160 × 19,323 nautical miles). The improved second stage can provide the thrust required to place a payload of 1,270 kilograms (2,800 pounds) into the same orbit. Both the standard and improved second stages have essentially the same components. The only difference is the size of the liquid-rocket engine and propellant tanks. Batteries charged prior to lift-off provide the electrical power required to perform the mission (see fig. 6-8).

In addition to the electrical system, the Delta Second Stage includes propulsion and guidance and control systems. The propulsion system for the standard second stage consists of a bipropellant

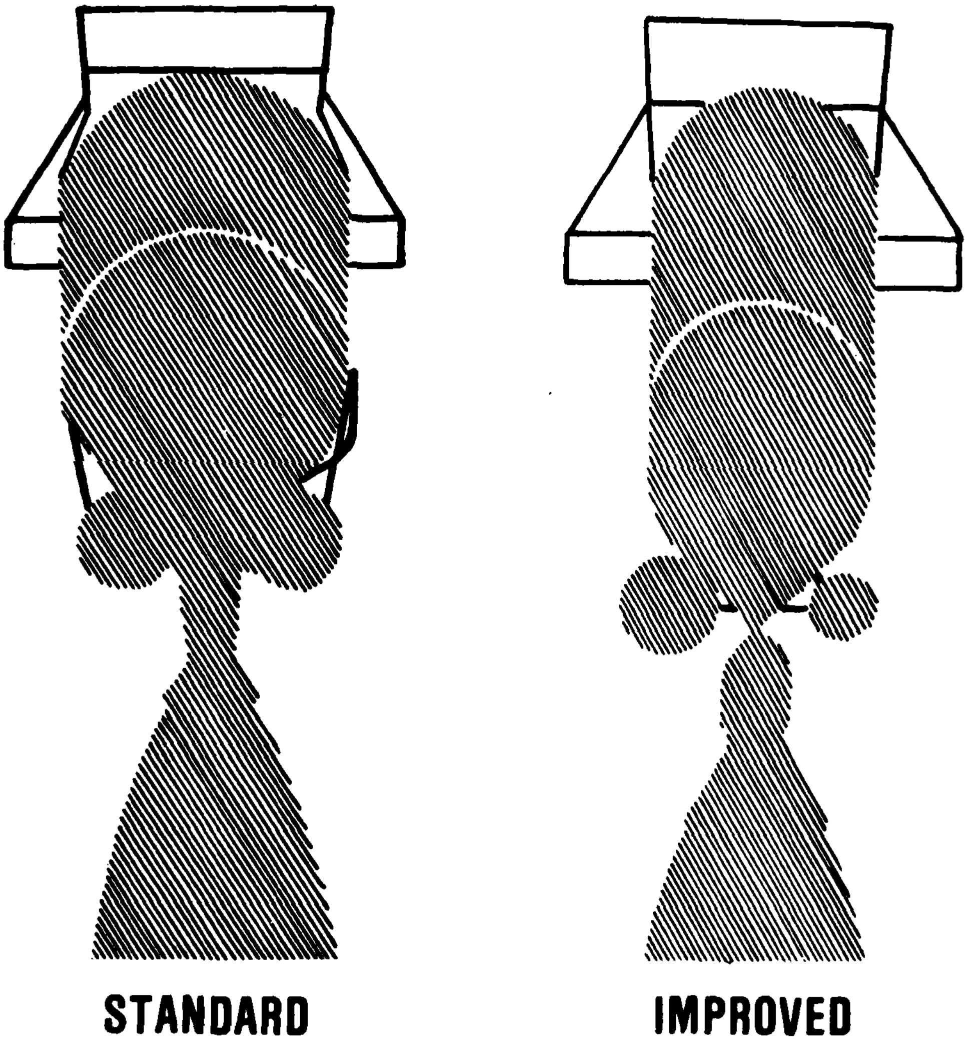


Figure 6-8 Delta Second Stage

liquid-rocket engine that produces thrust by the chemical reaction caused by combining two liquids in the engine's thrust chamber. The engine is capable of multiple starts. It produces 41.9 kilonewtons (9,425 pounds) of thrust over a total operating time of 322 seconds. It carries 4,536 kilograms (10,000 pounds) of propellant in the two storage tanks as well. In the improved second stage, the expansion ratio of the engine is increased, providing for more efficient performance, increasing the capability of the storage tanks by 1,451 kilograms (3,200 pounds), and increasing the total operating time to 444 seconds. Although the changes increase the thrust of the improved second stage by only 0.1 kilonewtons (20 pounds), the changes significantly increase overall performance as reflected in the payload weight-to-orbit numbers. The inertial guidance system provides the information for both first and second stage control. The Delta Second Stage issues commands to the Delta First Stage to gimbal, or move, the main engine in the pitch and yaw directions. Control of the second stage is much the same.

Transtage

The Titan IIIC used the Transtage from 1965 until 1982. Currently, the Titan 34D uses it as an upper stage. It consists of a control module and a propulsion module. It can produce the thrust required to place a payload of 1,859 kilograms (4,100 pounds) into a circular orbit of 35,786 kilometers (19,323 nautical miles). See figure 6-9.

The control module uses an inertial guidance section to deliver position and velocity data to the digital guidance computer. Then the computer issues the commands required to control both the Titan launch vehicle and the Transtage. These commands include the information necessary to jettison the solid-propellant motors strapped to the Titan, terminate the Titan thrust, and separate the Transtage and payload from the Titan. The module issues the commands required to control the firing of the propulsion module. Also it provides the electrical energy required by the guidance and control section and the propulsion module.

The propulsion module contains two propellant tanks, a pressurization unit, the structure to support the equipment, and two hydraulically-controlled liquid-rocket engines. The fuel is fed to the engines by opening valves that allow the fuel to enter the thrust chamber. The engines have an unlimited restart capability and produce 71.7 kilonewtons (16,000 pounds) of thrust over a total operating time of 440 seconds.

Payload Assist Module

Either the Delta launch vehicle or the Space Transportation System (STS) can use the Payload Assist Module (PAM). When used with these two, the PAM has similar mechanical interfaces and functional operations. The module consists of the payload attach fitting, STAR-48 solid-rocket motor, spin table, cradle, and airborne support equipment (see fig. 6-10).

The payload attach fitting provides the means of attaching the payload to the PAM solid-rocket motor, and it is the mounting location for the PAM subsystems. The engineers bolt the STAR-48 solid-rocket motor to the base of the payload attach fitting. It is the propulsive element of the module. Off-loading, or putting in less solid propellant during manufacture, adjusts the motor for a particular mission. The spin table provides rotational velocity to the PAM/payload combination for stabilization during nonpowered portions of the flight and during the solid-rocket motor burn. The cradle, used only on STS missions, mounts in the cargo bay of the orbiter and supports both the PAM and the payload. Since the airborne support equipment provides the interface with the STS electrical and safety systems, it is used only on STS missions.

When the Delta uses it, the PAM provides the thrust required to place a payload of 1,270 kilograms (2,800 pounds) into an orbit of 296 × 35,786 kilometers (160 × 19,323 nautical miles). When used with the STS, it provides the thrust required to place a payload of 1,052 kilograms (2,320 pounds) into an orbit 296 × 35,786 kilometers (160 × 19,323 nautical miles) from the nominal STS orbit of 296 kilometers (160 nautical miles).

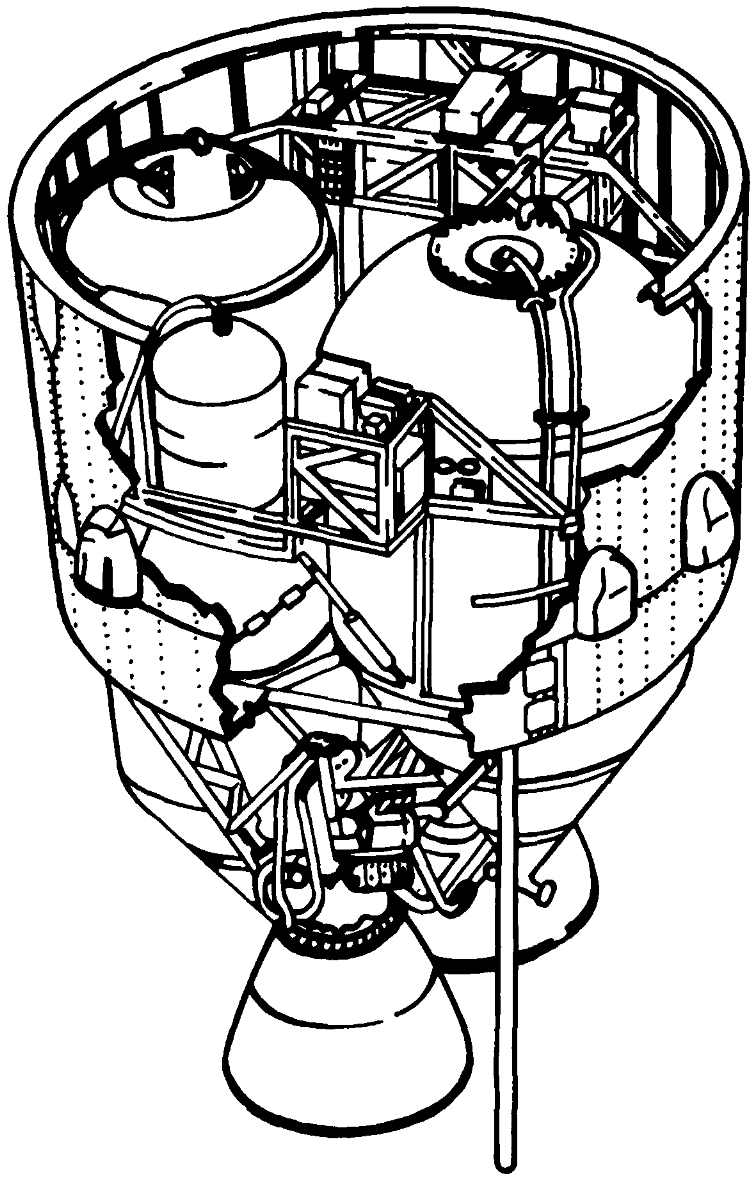


Figure 6-9. Transtage.

Inertial Upper Stage

The Inertial Upper Stage (IUS) is compatible with both the Titan 34D and the STS. It has two design configurations, namely, the Titan two-stage and the STS two-stage (see fig. 6-11). All configurations use almost identical active components; they vary only in their connective structures. The two-stage vehicle consists of the forward cone section, Stage II solid-rocket motor, an interstage, Stage I solid-rocket motor, and a rear staging section. The forward cone section provides avionics housing, payload attachment interface, and an equipment bay. The avionics consist of an inertial guidance system and a control system to command and control the vehicle and the payload after deployment from the STS. If a Titan is the launch vehicle, this system can control the entire mission.

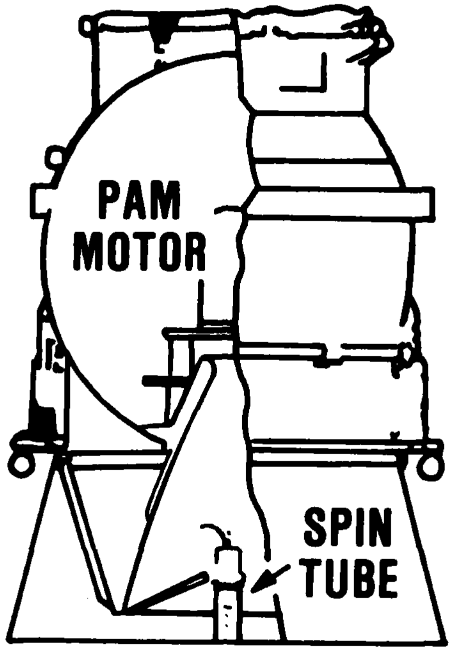


Figure 6-10. Payload Assist Module.

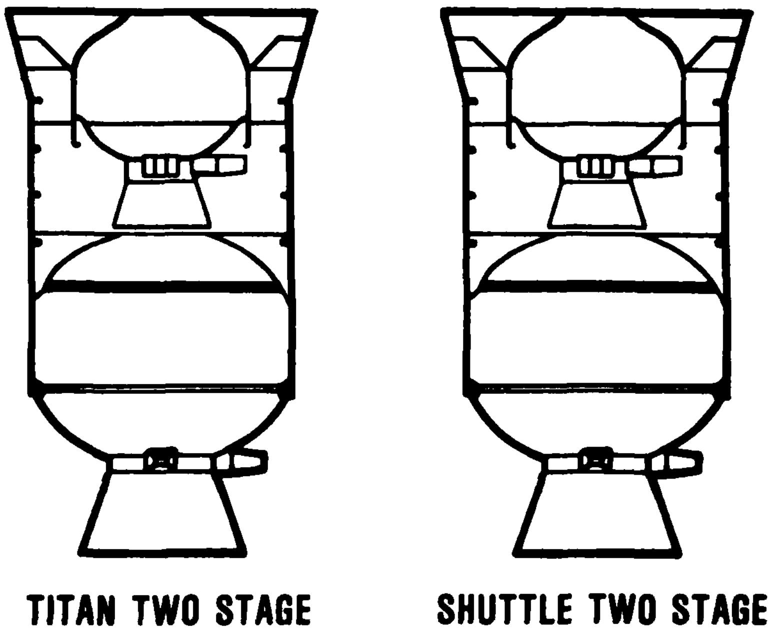


Figure 6-11. Inertial Upper Stage

The payload attachment interface, the same for either a Titan or STS mission, provides the mounting surface for attaching the payload. The equipment bay houses the telemetry and tracking systems and the batteries that supply electrical power.

The propulsion system consists of a large solid-rocket motor in the rear and a similar but smaller solid-rocket motor in the forward position. Each solid-rocket motor contains electromechanical actuators to move the nozzle for control of the vehicle during powered flight. For Titan missions, the engineers can off-load each motor during the manufacturing process to ensure the correct amount of propellant for a particular mission. A reaction control system provides vehicle control during unpowered portions of the flight. On Titan flights, the rear staging section provides the electrical and mechanical interface with the Titan launch vehicle. On STS flights, the rear staging section provides the mechanical interface to the cradle mounted in the orbiter bay and the airborne support equipment provides the electrical interface.

When the Titan 34D uses the IUS, it can provide the thrust required to place a payload of 1,859 kilograms (4,100 pounds) into an orbit 296 × 35,786 kilometers (160 × 19,323 nautical miles). When the STS uses it, the inertial upper stage can place a payload of 2,267 kilograms (5,000 pounds) into an orbit of 35,786 kilometers (19,323 nautical miles) from the nominal STS orbit of 296 kilometers (160 nautical miles).

Summary

The space launch vehicles described in this chapter are not compatible with all the upper stages (see table 6-1). However, the combinations of launch vehicles and upper stages available give the payload planner a wide-choice of payload capabilities (see table 6-2). The STS will replace the vehicles described in this chapter eventually because of its greater weight-to-orbit capability. However, these vehicles have served as the launch vehicles for the United States since the beginning of the space program in the 1960s and will continue to be used until the end of the 1980s.

BIBLIOGRAPHY

Aerospace Industries Association of America, Inc. *Aerospace Facts and Figures 1982-83.* New York: McGraw-Hill, 1982.

Boeing Aerospace Company. *Inertial Upper Stage User's Guide.* Seattle, Wash.: Boeing Aerospace Company. 1980.

General Dynamics Convair Division. *Centaur Mission Planners Guide.* San Diego, Calif.: General Dynamics Corp., 1977.

General Dynamics Convair Division. *General Dynamics Convair Division Space Transportation Systems.* San Diego, Calif.: General Dynamics Corp., 1982.

Martin Marietta Denver Aerospace. *Titan Launch Vehicles.* Denver, Colo.: Martin Marietta Denver Aerospace Corp., 1982.

Martin Marietta Denver Aerospace, *Titan 34B/34D Spacecraft User's Guide.* Denver, Colo.: Martin Marietta Denver Aerospace Corp., 1982.

McDonnell Douglas Astronautics Company. *PAM-D User's Requirements Document.* Huntington Beach, Calif.: McDonnell Douglas Astronautics Co., 1982.

McDonnell Douglas Astronautics Company, *The Delta Launch Vehicle: Past, Present, and Future.* Huntington Beach, Calif.: McDonnell Douglas Astronautics Co., 1981.

Table 6-1

Upper Stage Compatibility

	TITAN COMPATIBLE		ATLAS COMPATIBLE	
	Transtage	*IUS*	*Ascent Agena*	*Centaur D-1*
Height, m(ft)	4.5(14.7)	5.0(16.4)	6.3(20.7)	9.2(30.0)
Dia. m(ft)	3.1(10.0)	2.9(9.5)	1.5(5.0)	3.1(10.0)
Thrust, KN(lb × 1,000)	7.2(16.0)		75.6(17.0)	133.4(30.0)
Stage 1		186.8(42.0)		
Stage 2		75.6(17.0)		

	DELTA COMPATIBLE		
	2nd Stage	*Improved 2nd Stage*	*PAM-D*
Height, m(ft)	60.0(19.6)	6.0(19.6)	3.5(11.6)
Dia. m (ft)	1.4(4.5)	1.8(5.7)	1.3(4.2)
Thrust, KN(lb × 1,000)	41.8(9.4)	41.8(9.4)	84.5(19.0)

	SHUTTLE COMPATIBLE		
	IUS		*PAM-D*
	Two Stage	*Twin Stage*	
Height, m(ft)	5.0(16.4)	7.5(24.4)	3.5(11.6)
Dia. m(ft)	2.9(9.5)	3.1(10.2)	1.3(4.2)
Thrust, KN(lb × 1,000)			
Stage 1	196.2(44.1)	196.2(44.1)	84.5(19.0)
Stage 2	74.7(16.8)	196.2(44.1)	

Table 6-2

Current US Space Launch Capabilities

VEHICLE	STAGES	THRUST kilonewtons (pounds × 1,000)	LAUNCH SITES*	ORBIT kilometers (nautical miles)	INCLINATION	PAYLOAD WEIGHT kilograms (pounds)
SCOUT	1. Algol III A 2. Castor II A 3. Antares III A 4. Altair III A	431.1(96.9) 285.2(64.1) 83.1(18.7) 25.6(5.8)	SLC-5 WSMC	556(300)	0°	181(400)
Delta 3914	0.9 Castor IVs 1. Extended Long Tank 2. Standard 2ND Stage 3. TE-M-364-4 Solid Rocket Motor	378.0(85.0)ea 921.0(207.0) 41.8(9.4) 65.8(14.8)	LC-17 A&B ESMC SLC-2W WSMC	185 × 35,786 (100 × 19,323)	27°	937(2,065)
Delta 3910/PAM	0.9 Castor IVs 1. Extended Long Tank 2. Standard 2ND Stage 3. Payload Assist Module (PAM)	378.0(85.0)ea 921.0(207.0) 41.8(9.4) 84.5(19.0)	LC-17 A&B ESMC SLC-2W WSMC	296 × 35,786 (160 × 19,323)	27°	1,111(2,450)
Delta 3920/PAM	0.9 Castor IVs 1. Extended Long Tank 2. Improved 2ND Stage 3. Payload Assist Module (PAM)	378.0(85.0)ea 921.0(207.0) 41.8(9.4) 84.5(19.0)	LC-17 A&B ESMC SLC-2W WSMC	296 × 35,786 (160 × 19,323)	27°	1,270(2,800)
Atlas E	1. MA-3 2-Boosters 1-Sustainer 2. SGS II	1,748.1(393.0) 65.8(14.8)	SLC-3 WSMC	157 × 20,198 (85 × 10,900)	63°	1,088(2,398)

Table 6-2 (Continued)

Current US Space Launch Capabilities

VEHICLE	STAGES	THRUST	LAUNCH SITES*	ORBIT	INCLINATION	PAYLOAD WEIGHT
		kilonewtons (pounds × 1,000)		kilometers (nautical miles)		kilograms (pounds)
Altas H	1. MA-5 2-Boosters 1-Sustainer 2. SGS II	1,948.2(438.0) 65.8(14.8)	LC-36 ESMC SLC-3 WSMC	157 × 20,198 (85 × 10,900)	55°	1,225(2,700)
Atlas SLV-3D Centaur	1. MA-5 2-Boosters 1-Sustainer 2. Centaur DI-A (RL-10)	1,912.6(430.0) 133.4(30.0)	LC-36 ESMC	185 × 35,786 (100 × 19,323)	27°	1,996(4,400)
Atlas G Centaur	1. MA-5 2-Boosters 1-Sustainer 2. Centaur (RL-10-3A)	1,948.2(438.0) 146.8(33.0)	LC-36 ESMC	185 × 35,786 (100 × 19,323)	27°	2,359(5,200)
Titan 34B Agena	1. LR-87 2. LR-91 3. Ascent Agena	2,059.4(463.0) 449.2(101.0) 75.6(17.0)	SLC-4 WSMC	185(100) Circular	90	3,583(7,900)
Titan 34D NUS	0. Two 5½ Segment SRMs 1. LR-87 2. LR-91	12,454.4(2,800.0) 2,353.0(529.0) 449.2(101.0)	SLC-4 WSMC	185(100) Circular	90	12,520(27,600)
Titan 34D Transtage	0. Two 5½ Segment SRMs 1. LR-87 2. LR-91 3. Transtage	12,454.4(2,800.0) 2,353.0(529.0) 449.2(101.0) 71.2(16.0)	LC-40 ESMC	35,876(19,323) Circular	0	843(1,859)

Table 6-2 (Continued)

Current US Space Launch Capabilities

VEHICLE	*STAGES*	*THRUST*	*LAUNCH SITES**	*ORBIT*	*INCLINATION*	*PAYLOAD WEIGHT*
		kilonewtons (pounds × 1,000)		kilometers (nautical miles)		kilograms (pounds)
Titan 34D IUS	0 Two 5½-segment SRMs	12,454.4 (2,800.0)	LC-40 ESMC	35,876 (19,323) Circular	0	843 (1,859)
	1 LR-87	2,353.0 (529.0)				
	2 LR-91	449.2 (101.0)				
	3 IUS					
	Stage 1	186.8 (42.0)				
	Stage 2	75.6 (17.0)				
STS	0 2 Solid Rocket Boosters	11,155.0 (2,508.0) ea	LC-39 ESMC	278 (150) Circular	28.5	29,500 (65,000)
	1 3 Main Engines	1,668.0 (375.0) ea				
	2 Inertial Upper Stage (Two Stage)	196.0 (44.1) 74.7 (16.8)	STS Parking orbit	35,786 (19,323) Circular	0	2,267 (5,000)
	or 2 Inertial Upper Stage (Twin Stage)	196.2 (44.1) 196.2 (44.1)	STS Parking orbit	Planetary Missions	N/A	4,993 (11,009)
	or 2 PAM-D	84.5 (19.0)	STS Parking orbit	278 × 35,786 (150 × 19,323)	27	1,052 (2,320)

*Launch sites at the Eastern Space and Missile Center (ESMC) are designated as launch complexes (LC) and if served by the same launch control center, are further designated as A or B.

Launch sites at the Western Space and Missile Center (WSMC) are designated as Space Launch complexes (SLC) and if served by the same launch control center are further designated as E(ast) or W(est).

Chapter 7

GLOBAL COMMUNICATIONS

Closely associated with the guidance and control system for a space vehicle is the communications system. Since the early part of the present century when Marconi developed his "wireless" telegraph, people have been sending messages through space. Today, the means exist already for communicating as far as the known limits of the Solar System. The problems that the space communications engineer must solve are not involved with the generation and propagation of electromagnetic waves, but rather with problems that arise because of the complexity of the system components and the nature of the host vehicle. Such problems are those involved in allowing for the cubage and mass of the equipment in the host vehicle and in generating electrical power in space.

COMMUNICATION SATELLITES

Satellite systems provide global command and control networks and permit the simultaneous transmission of a greater number and variety of high-quality messages. In addition, there is a higher degree of reliable transmission and a greater degree of survivability designed into the system.

In this age of jet aircraft, nuclear power, and moon rockets, present communication facilities simply have become outmoded. These facilities have served well and will continue to do so in the future, but they are not good enough. They are unable to handle the greatly increased volume of communications. Present communications are disrupted easily. A slight increase in solar flare activity can cause radio communication blackout between Europe and North America. A fishing trawler may cut accidentally the North Atlantic telephone cable. Although these events present difficult problems to civilian communications, the implications of even temporary communication isolation from the rest of the world are far more serious for a military commander.

Objectives

What should be some of the major objectives in developing a system of communication satellites? One objective is reliable communications that provide uninterrupted service over long periods. The system should have high capacity with capability for handling large volumes of all types of traffic. It should be flexible to serve the maximum number of potential users. There should be minimum delay in transmission. We will discuss each of these factors in more detail, and compare the communication satellite with some existing facilities.

Reliability

Two types of reliability are of interest. The first is propagation reliability. The high-frequency (HF) band has always been subject to the vagaries of the ionospheric layers that surround the earth. Thus, only a portion of this band actually is usable at any given time over a particular path. In addition, multipath effects seriously limit the amount of information that we can transmit over a given channel. Added to these limitations are the blackouts that may result from ionospheric disturbances caused by sunspot activity. High-altitude nuclear explosions can introduce similar disturbances. We are forced

to the conclusion that HF radio via the ionosphere is less than satisfactory as a propagation medium.

By contrast, a communication satellite of the active repeater type employing line-of-sight transmission at microwave frequencies would be extremely reliable from a propagation standpoint. However, the communication satellite introduces a second type of reliability problem, that of reliable unattended operation for long periods in orbit. Scientists have demonstrated that someone can develop a reliable communication satellite with lifetimes of many years if the individual uses the following practices: selecting components of proven reliability; operating all components well within their ratings; providing adequate protection by satellite design during launch and in the space environment; and, finally, using redundancy adequately to further increase the probability of successful operation.

High Capacity

Through the years people have been dependent, almost exclusively, on high frequencies in the band 5 to 30 megahertz* for long-range global communications. All countries share these frequencies. They must support both military and civilian applications. The narrow range of frequencies, and the propagation characteristics discussed previously, seriously limit the total communication capacity.

It is not surprising that great interest continually exists in new techniques that promise to open additional areas of the frequency spectrum to long-range communications. Examples are ionospheric and tropospheric scatter propagation. The communication satellite has opened the complete range of frequencies to 10,000 megahertz and beyond for long-range communications, thus providing nearly 1,000 times the spectrum available in the HF band.

Flexibility

One requirement is to provide sufficient flexibility in a system so that it can satisfy new or changing demands around the world without major overhaul or replacement of facilities. For instance, a disadvantage of the submarine cable is lack of flexibility with a point-to-point, fixed-plant facility. On the other hand, by providing wide bandwidths and essentially global coverage, a system of communication satellites in 24-hour equatorial orbits places minimum restraint on the number and location of ground stations served and the volume of communications furnished to each.

Minimum Delay

Another objective is to speed up communications. All too frequently, congested facilities cause urgent message delays because propagation conditions are poor. Some of the advantages of communication satellites, which we discussed earlier, such as the bandwidth, not only make possible worldwide television broadcasting in real time, but also reduce delays in all types of communications.

SURVEY OF TYPES OF COMMUNICATION SATELLITES

Thorough review of all factors has led to the conclusion that an active repeater satellite in a 24-hour equatorial orbit offers the most promise for advancing global communications. However, in view of the reliability problems and the anticipated costs of a communication satellite program, we should review briefly the pros and cons of proposed alternative approaches.

Passive Reflector

The passive system uses a reflecting surface that cannot amplify or retransmit signals. Some of the advantages of a passive system are its inherent reliability and the possibility of a large number of users sharing it as it operates over a wide range of frequencies. However, operating between two locations 2,000 miles apart, a typical system would require 24 balloons 100 feet in diameter in randomly-spaced

*Hertz the unit of frequency cycles per second KHz (kilohertz) is 1,000 cycles per second MHz (megahertz) is 1,000,000 cycles per second

orbits at 3,000 miles altitude for an outage time of 1 percent. The system would require substantially more satellites to provide longer range or wider coverage than the sample cited. Therefore, such satellites do not appear to be an economical solution to providing truly global communications.

The passive reflector satellites would be more attractive if they could be stabilized in attitude, permitting use of more efficient reflecting surfaces, or could synchronize the satellite orbits to reduce the required number of satellites. However, then the satellites cease to be passive and reliability is no longer inherent in the system.

Active Repeater

Active communications satellites receive signals, translate them in frequency, and amplify and retransmit the signal at a higher power level. We can consider them repeater stations in space. The use of active communications satellites makes possible the use of smaller ground terminals. This enhances flexibility of military operations. Because of the increased radiated power over that of the passive reflector, the transmission path loss is not as great a problem. Therefore, we can place the active satellites in orbit at much higher altitudes. Because the number of satellites required to provide continuous coverage for a number of ground terminals varies as the orbit altitude changes, the higher we place the system, the fewer number of satellites it requires to provide this continuous coverage.

Medium-Altitude System

Continuous coverage from medium altitude would require from 18 to 24 satellites in orbit. Even then, there would be a switching problem at the ground terminal as one satellite passed from view and a new one approached to take its place. The ground terminals require steerable antennas as well as computing equipment to calculate the trajectories and furnish look-angles (acquisition data) for antenna orientation.

A major advantage of the medium altitude system is that it simplifies orbit injection and does not require precise position stabilization. The booster requirements are not as great either.

Synchronous-Altitude System

Because of the earth's rotation, three satellites at an altitude of 19,360 nautical miles and equally spaced in 24-hour equatorial orbit appear to remain fixed to an observer on the earth. The high altitude of the synchronous orbit (fig. 7-1) makes each satellite visible from 40 percent of the earth's surface. Since the satellite appears to be motionless in the sky, service may begin or continue with only one satellite in orbit and functioning. However, in actual operation, we would need four or six operating satellites to provide the desired coverage. Although this number is less than the requirement for a medium altitude system, the synchronous altitude system has a much greater booster requirement. The reason for the increased booster requirement is the necessity to interject approximately the same amount of satellite weight at a much greater altitude.

CONSTRAINTS

To appreciate the capabilities of space communications more fully, we need to consider the natural constraints and design limitations.

Natural Constraints

The natural constraints include line of sight, space attenuation, and noise. To operate, there must be a line of sight established between the transmitter and the receiver. The transmitting antenna radiates power that is distributed over an always expanding portion of a spherical surface. The resulting decrease in power density (power per unit area) reduces the energy that the receiving antenna captures. We know this as space attenuation. Various sources introduce noise at each stage of the communications process. The communications mediums used to send the signals and the receiver

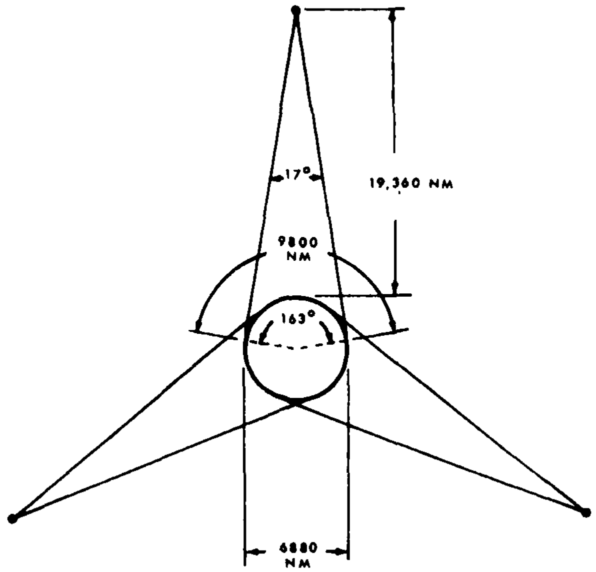


Figure 7-1. Concept of three communication satellites in 24-hour equatorial orbit, showing geometrical relationships in the equatorial plane.

itself are the most significant contributors. Noise reduces the ability of the receiver to detect weak signals.

Design Limitations

These include transmitter power, receiver sensitivity, receiver noise figure, and bandwidth. Obviously, power must be available at the transmitter for operation. Radio sensitivity is a measure of the minimum signal strength with which we can operate the receiver gainfully. The receiver's injection of noise into the system constitutes a basic limit on the minimum detectable signal. Many considerations limit the bandwidth of the system. The most important of these is that the system's capacity to transmit data is directly proportional to its usable bandwidth.

Data Processing and Modulation

The final two design limitations we will mention are data processing and modulation. Only the transmission of information constitutes a profitable expenditure of energy for a communications system. All data does not constitute information. The effective capacity of a system is directly proportional to the efficiency with which the transmitted data represents information. The effectiveness of the communications system varies greatly with the modulation technique used. (Modulation is the process of imposing signal data on a carrier much as the lips and tongue "modulate" the "carrier" generated by the larnyx.) We will explain the most important constraints on the communications facility in more detail.

Line-of-Sight Transmission

A prime limitation on the communications facility is the necessity of establishing line-of-sight transmission between the transmitting and receiving antennas. This applies in every case. We can obtain virtual line of sight when direct line of sight is not possible.

Figure 7-2 shows direct line-of-sight coverage from a transmitter. The receiver is below the radio horizon represented by the dotted line (located where a cone with the apex at the transmitter would be

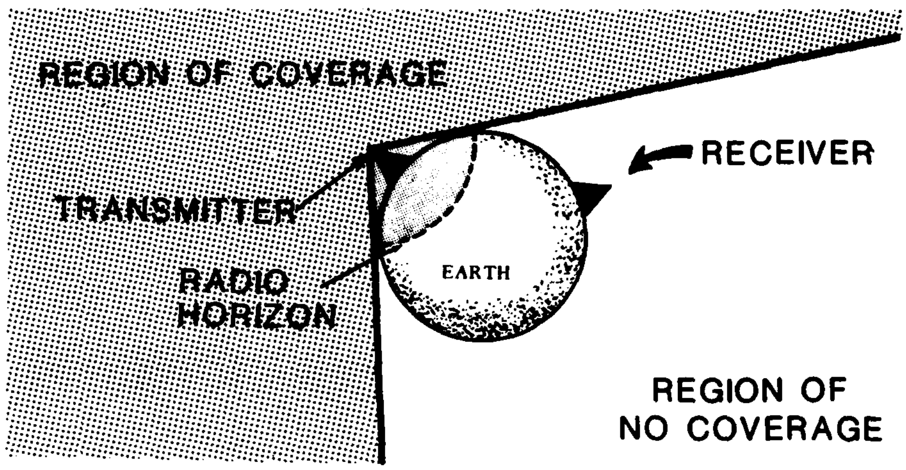


Figure 7-2. Direct line-of-sight coverage of a transmitter.

tangent to the earth's surface). The question naturally arises as to the possible extension of the coverage. After all, we receive radio transmissions at points far below the radio horizon. It is easy to understand why. The near-earth environment is not empty space. It is full of oxygen, nitrogen, water vapor, charged particles, magnetic fields, and other material. Of particular interest is the ionosphere, which consists of layers of charged particles that have the property of transmission, refraction, or reflection of radio waves, depending on the properties of the layer, the geometry of transmission, and the frequency involved. Below certain critical frequencies, or maximum usable frequencies (MUF), the charged layers of the ionosphere affect radio waves in much the same manner that a partial mirror may reflect and refract light waves (see fig. 7-3).

In figure 7-3, the ionosphere reflects the transmission from A to the remote position B, which lies below the radio horizon of the transmitter. As figure 7-4 shows, the receiver does not actually see the location of the transmitter as being at point A. As far as the receiver at B is concerned, the transmission originates at point A′. The ionospheric mirror establishes the virtual image of the

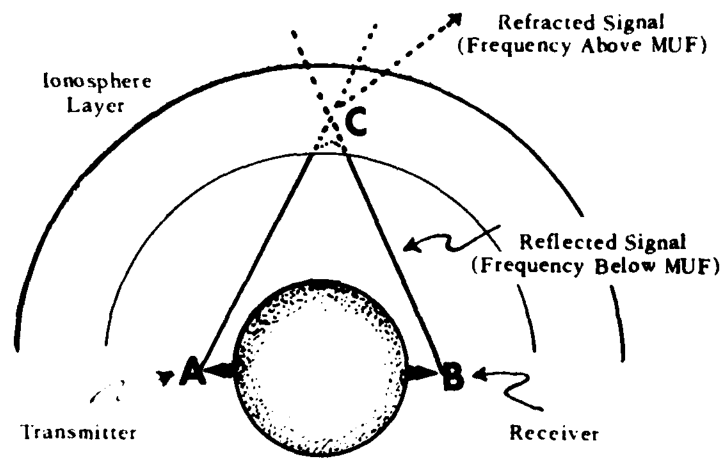


Figure 7-3. Reflection and refraction of radio waves by the ionosphere.

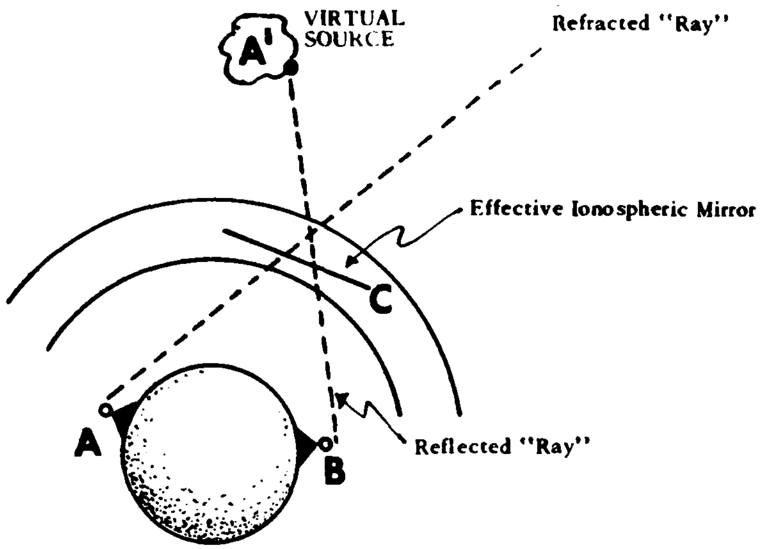


Figure 7-4. Virtual image of a transmitter established by reflection from the ionosphere.

physical transmitter at point C'. Thus, we have established a line of sight between a virtual source at A' and the receiver at B. We have fulfilled the requirement of line-of-sight transmission from the transmitter to the receiver.

This is analogous to seeing around a corner by means of a mirror as figure 7-5 shows. With the mirror arranged as shown, we do not see the object at its actual position but at a virtual position that optical geometry determines. Remember in using this analogy that, in the case of the ionosphere, the virtual source is not so well-defined as the actual source is. The virtual source has become enlarged. The turbulent condition of the ionosphere has blurred its boundaries.

This "method of mirrors" seems to solve the problem in a simple way. Nature furnished the reflecting medium, so why look for a better scheme? However, we need something better. Being in a constant state of flux, the ionosphere is unreliable as a transmission medium. The properties of the ionosphere vary from year to year, from day to day, and even from hour to hour (see fig. 7-6). Also, energy in the maximum usable frequencies reflected from the ionosphere is too low to provide the desired information-carrying capacity on a channel using three frequencies. (The higher the carrier frequency, the higher the data capacity of the channel.) We desire use for carrier frequencies of an order-of-magnitude higher than the MUF of ionospheric reflection technique permit.

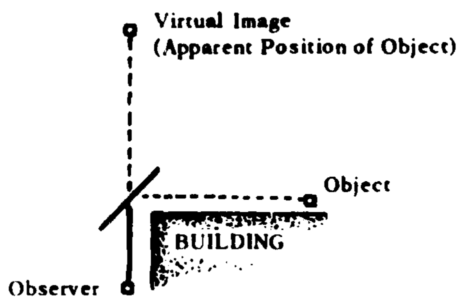


Figure 7-5. Virtual image of an object formed by reflection in a mirror.

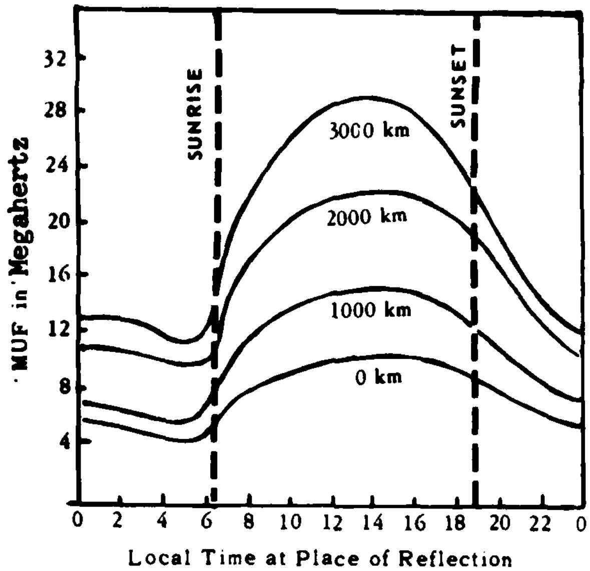


Figure 7-6. Variations in the ionosphere as reflected by hourly changes in the maximum useable frequency.

Another important problem arises because the ionosphere is distributed widely spatially and it has considerable thickness. Therefore, the transmitter and the receiver have several transmission paths of varying lengths established between them. The result of this multipath transmission is fading, garbling, and marked reduction in the effectiveness of transmission.

To use a virtual source to see around the curve of the earth, we must fix the properties of the system so that they are predictable, constant, and localized in space. Theoretically, one way to do this would be to construct tall antennas, as figure 7-7 shows.

However, someone calculated that antennas for transatlantic communications would have to be 360 miles high. To construct such towers would take the gross national product of the United States at the current rate for thousands of years.

A more practical way would be to place a satellite at the vertical source A′, as figure 7-4 shows. The satellite serves as a controlled, predictable, space-localized transmitter (or passive reflector) at the virtual source. By selecting frequencies that are essentially independent of the ionosphere, the system should have the capacity, reliability, and coverage to please the most demanding commander.

Space Attenuation

When we remove the receiver far from the transmitter, the signal is weaker than when the receiver is near the transmitter. One of the reasons for this is a purely geometrical relationship known as space

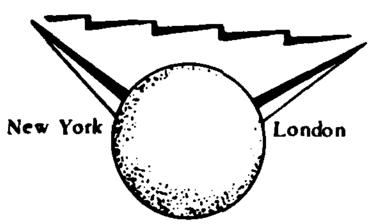


Figure 7.7 Tall antennas used to overcome the curvature of the earth.

attenuation. As the term implies, the signal strength decreases as the intervening transmission path becomes longer. However, this decrease is not a loss in the sense of an irreversible conversion of electromagnetic energy to some other energy form, such as heat.

Attenuation of the signal constitutes one of the basic limitations on communication over extensive distance. Space attenuation is not a loss. It occurs because the receiving antenna is of limited physical dimensions. If we could construct an antenna to enclose the source completely, space attentuation would not occur.

Scientists employ many techniques to minimize the effects of space attenuation. In general, these techniques recognize that isotropic radiation is seldom a necessary specification for a given system. Transmitted power does not have to travel uniformly in all directions from the source. Instead, we can focus the power to travel in the direction preferred. We call an antenna that can focus power a directional antenna. Perhaps the most familiar example of a directional antenna is the parabolic radar antenna. This antenna serves a dual purpose, namely, to form a narrow beam and to increase greatly the power density within the beam over that accomplished by radiation from an isotropic source. Even in interplanetary communications, isotropic radiation would not be necessary since the orbits of the planets are roughly in a common plane. Thus, transmissions from a point within the Solar System would use a pancake-shaped pattern. This would have greatly enhanced capability over an isotropic radiator (see fig. 7-8).

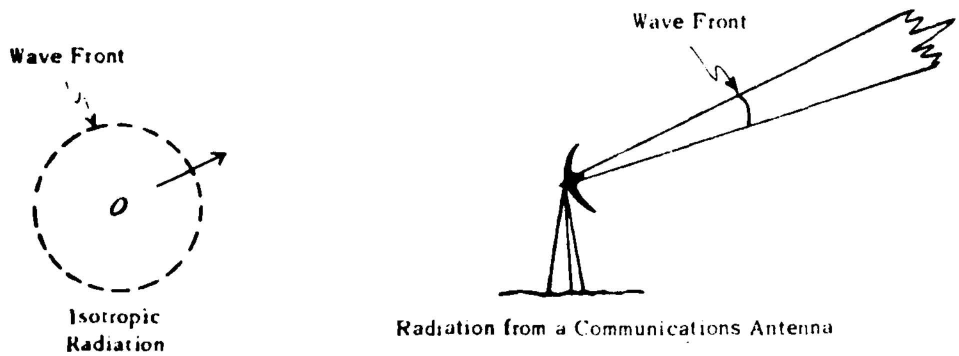


Figure 7-8. Comparison of isotropic radiation with directional radiation from a communications antenna.

Noise

Most people are familiar with the crackling static that plays havoc with programs on the broadcast band during electrical storms, particularly if the station is not strong. Indeed, during heavy electrical disturbances we may have to stay close to the receiver to hear. In this case, natural interference such as noise injected into the system through the antenna has curtailed severely the communications range.

Often, the generation of noise occurs within the receiver itself. If we tune an ordinary radio to a frequency with no incoming signal, and we increase the volume, we can hear a hissing background. The source of this noise is within the components and we cannot eliminate it. However, by careful design, we can reduce this source of noise to a very low level.

We need to consider the question of which noise source gives more trouble. Natural interference predominates in the lower part of the radio frequency spectrum. This is in the low- and high-frequency bands or in frequencies up to approximately 30 megahertz. At frequencies above this, the noise generated within the receiver itself predominates to the extent that it can neglect external natural noise. During an electrical storm a radio may be inaudible, but the television may not be affected.

This happens because radio operates at frequencies far below 30 megahertz (in the region of natural noise), and television operates above 30 megahertz where natural interference does not present a major problem. Since the maximum usable frequency for transmission by ionospheric reflection is below 30 megahertz, natural interference is present when we use ionospheric transmission (see fig. 7-9).

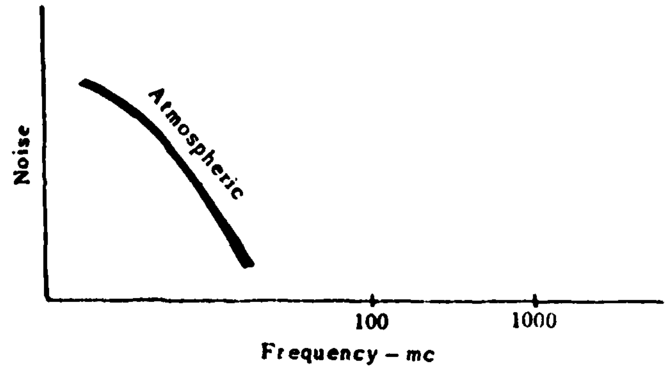


Figure 7-9. Variation in noise level with the frequency of the radio band.

Modulation

Modulation is the process that we use to impart intelligence or signal data to the radio wave or carrier. A radio signal is not a simple, single-frequency (monochromatic), sinusoidal oscillation that generally we associate with wave motion. Rather, it is a carrier wave varied by some feature, such as amplitude or frequency, that varies with the intelligence (data) being transmitted. The name for this variation of some feature of the radio wave is modulation. The fundamental radio frequency upon which the modulation is imposed is the carrier. The carrier frequency of a radio signal is the frequency, or number on the dial to which we tune the receiver to receive a particular signal. We can represent the two most common modulation techniques, amplitude modulation and frequency modulation, graphically as figure 7-10 shows.

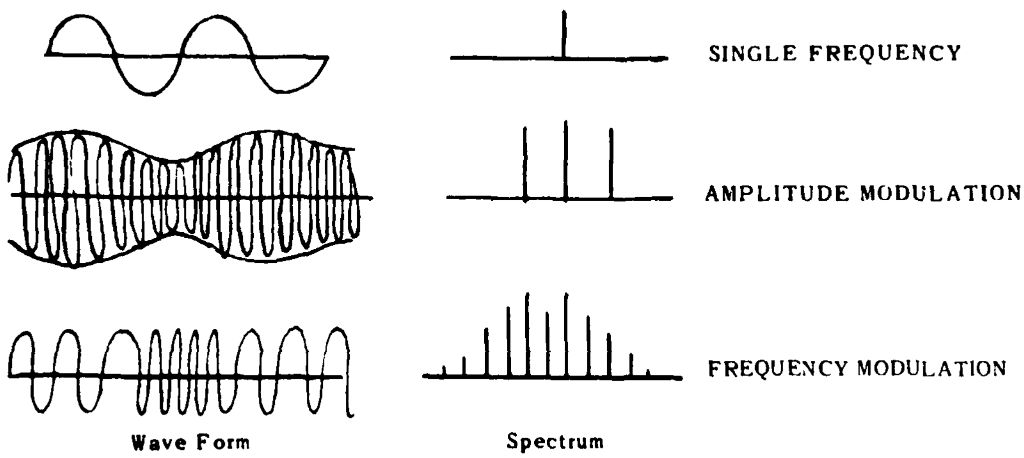


Figure 7-10. An amplitude-modulated radio wave and a frequency-modulated radio wave compared with a single-frequency wave.

Bandwidth

The name for the band of frequencies which the signal occupies is the signal bandwidth. We refer to the components of the signal above and below the carrier as the signal sidebands. It is apparent that the receiver window, or bandwidth, must be at least as large as the signal itself to receive the entire signal. It is interesting to note that the carrier and sideband components of a radio signal are not a mathematical fiction, but we can observe them with proper equipment such as a spectrum analyzer. The significance of the sideband components lies in the fact that they contain the data or intelligence of the signal. Transmission from the transmitter must take place, followed by propagation through space, and, finally, reception and reproduction at the receiver, equally well. Otherwise, we will lose part of the intelligence signal. Consequently, the data rate, or capacity, of a communications system largely determines not only its operational characteristics but also its technical requirements. Examples of bandwidths required for common types of communications service are 100 hertz for teletype, a low capacity or low data-rate system; 3,000 to 5,000 hertz for voice communications; and 6,000,000 hertz for commercial television, which is a high-capacity or high-data-rate system containing both aural and visual detail.

As figure 7-10 shows, the modulated ratio wave is a sinusoidal wave that we have distorted in either amplitude or frequency, or in some other manner, to provide intelligence. When only the carrier frequency is present, the bandwidth can be, theoretically, infinitesimally narrow, to encompass only that particular frequency. However, when we modulate the carrier, the bandwidth must encompass the carrier frequency and the frequencies superimposed upon it as sidebands. The total number of frequencies represented in the modulated wave is a function of the rate at which the modulation and variance of the wave takes place. For example, a radio wave with a carrier frequency of 10 kilohertz (KHz) that a 1,000-hertz tone modulates will contain frequency components from 9 KHz to 11 KHz. Therefore, when considering the normal commercial broadcasting stations, the bandwidth to pass such a radio wave must be 2 KHz in width. Special types of transmission, such as single sideband, do not require as large a bandwidth. But, in standard practice, the bandwidth must be at least as large as the highest modulating frequency, and in many applications very much larger.

Earlier, we indicated that noise determines the minimum usable signal power level. This includes noise entering the system with the signal at the antenna, and noise generated in the receiver itself. Generally, we consider electrical noise to be Gaussian in nature and widely distributed throughout the spectrum. Although desirable from a data-rate and capacity viewpoint, wide bandwidths are undesirable from the viewpoint of noise. The wider the bandwidth, the greater is the average noise power that enters the system, both at the antenna and within the electrical circuits. We could raise the power transmitted to restore the signal to an acceptable minimum signal-to-noise ratio, but power is not always easily attainable. We must consider the additional factors of the relation of power to bandwidth. The total power in the signal consists of the power in the carrier frequency and that of the sideband frequencies. In an amplitude-modulated signal as much as one-third of the total signal power resides in the sidebands, and we must transmit the entire bandwidth or we will lose part of the intelligence. Quite obviously, wide bandwidth, which high data rates and high capacity require, needs high power to maintain a usable signal-to-noise ratio. In summary, the high data rates require large bandwidths and high power and the average noise power is proportional to bandwidth.

MILITARY COMMUNICATIONS*

Reliable, survivable, and secure command, control, and communications (C^3) systems enhance flexibility in employment of our diverse military focus. The military has the C^3 systems interlaced throughout the total DOD structure, including the land, sea, and aerospace forces and their supporting elements. Effective C^3 requires that all of the individual parts be coordinated so that the appropriate mix of military power can be brought to bear on a crisis or war situation in precise

*Extracted from "C^3, Today and Tomorrow," by Dr Malcolm R. Currie, *Commanders Digest* (May 15, 1975)

response to and under the positive control of the proper national authorities. These C^3 systems should be interoperable and capable of providing rapid and secure communications between a variety of levels of command, between forces of the various services, and between the United States and its allies.

The ability to exercise command and control of our forces has become increasingly significant during recent years due to the increasing complexity of the tasks facing us, coupled with decreasing force levels.

The DOD is acquiring and deploying communications satellite systems that primarily address three fundamental communication areas. These are:

- C^3 of nuclear-capable forces for which a general-war survivable system is required. AFSATCOM fulfills this requirement.

- Support of the Worldwide Military Command and Control System (WWMCCS) using a day-to-day global system that provides protected strategic connectivity between major bases and command centers, large Navy ships, and advanced airborne command posts. This is the prime mission of the Defense Satellite Communications System (DSCS).

- Fleet communications modernization using a global system to serve the beyond-the-horizon communication needs of ships, antisubmarine warfare aircraft, and other mobile forces. This is FLTSATCOM. For more information on these systems, see chapter 8.

COMMERCIAL COMMUNICATIONS

In July 1963, NASA launched the world's first synchronous communications (SYNCOM) satellite into orbit. The purpose of NASA's SYNCOM program was to demonstrate the feasibility of synchronous orbit satellite communications. Since then, there have been more than 31 commercial satellites put into orbit. A very large percentage of these are either operating or on an orbital reserve. We will discuss six major systems in operation today.

INTELSAT IV

First launched in 1971, seven of these large high-capacity communications satellites are on orbit. The satellites have steerable antennas that can be aimed to cover selected areas on earth. Each satellite has an average capacity of 6,000 telephone channels or 12 simultaneous color television channels, or various combinations of telecommunications traffic including DATA, TELEX, and FACSIMILE (see fig. 7-11).

WESTAR

NASA launched America's first domestic communications satellite for Western Union on 13 April 1974. NASA launched the fifth in the series in 1982. Scientists designed the satellites to relay voice, video, and data communications to the continental United States and Alaska, Hawaii, and Puerto Rico. Each satellite has a capacity of 7,000 two-way voice circuits or 12 simultaneous color TV channels. On-orbit design lifetime for each satellite is seven years (see fig. 7-12).

INTELSAT IV-A

September 1975 was the date of the launching of the first of a new series of INTELSAT satellites. It had nearly twice the communications capability of its predecessor. A spot beam antenna system on the IV-A satellites permits reuse of the same frequency by spatial separation of the beams. This separation allows use of one set of frequencies in North and South America and reuse of the same frequencies in Europe and Africa. Each satellite has capacity for 11,000 voice channels or 20 simultaneous color TV channels (see fig. 7-13).

Figure 7-11. INTELSAT IV.

RCA-SATCOM

The RCA-SATCOM system consists of three satellites in geostationary orbit to serve the contiguous United States and Alaska with TV, voice channels, and high-speed data transmission. The three launchings occurred in December 1975, March 1976, and November 1981 (see fig. 7-14).

Figure 7-12. WESTAR.

MARISAT

Designed to provide reliable high-quality communications for the US Navy and commercial shipping, MARISAT satellites are now in orbit over the Indian, Atlantic, and Pacific Oceans.

These specialized spacecraft, developed for COMSAT General Corporation and a consortium of International Telephone and Telegraph, Radio Corporation of America, and Western Union, provide voice, teletype, facsimile, and data services. The Navy leased half the capacity of this system as an interim measure until FLTSATCOM became available in February 1978 (see fig. 7-15).

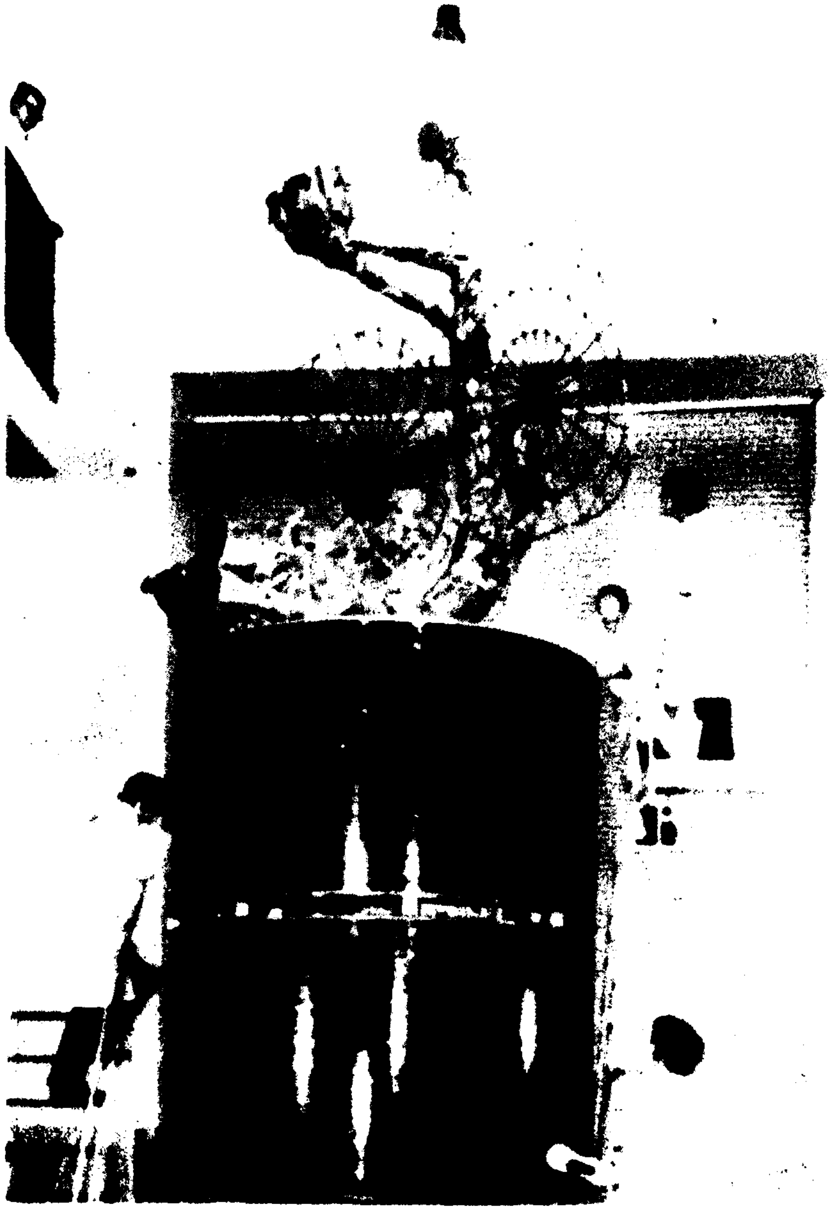


Figure 7-13. INTELSAT IV-A.

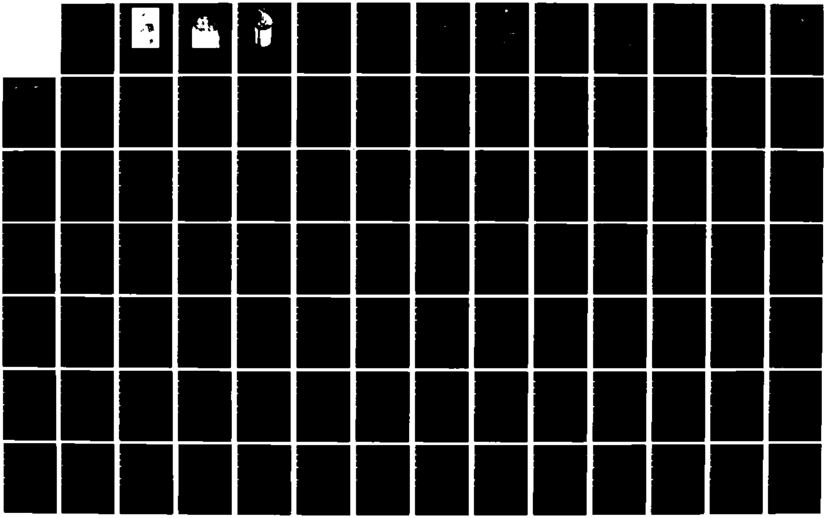
COMSTAR

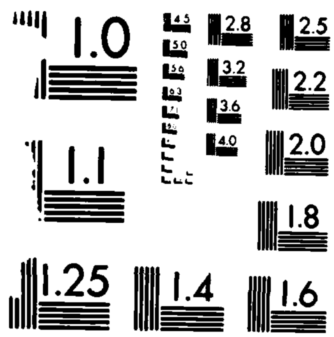
The COMSTAR satellites, launched for the COMSAT General Corporation and then leased to American Telephone and Telegraph have a capacity of 14,000 high-quality voice circuits. The system of four geostationary satellites provides service to the continental United States, Hawaii, Puerto Rico, the Virgin Islands, and Alaska (see fig. 7-16).

Commercial communications satellites provide redundant circuits for DOD communications worldwide. Commercial satellites provide a backup for the DOD communications satellites. If a primary DOD communications satellite fails, the DOD can route critical information through the commercial satellites.

AD-A165 038 SPACE HANDBOOK(U) AIR MINISTRY LONDON (ENGLAND) 3/4
C D COCHRAN ET AL. JAN 85 AU-18

UNCLASSIFIED F/G 15/7 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Figure 7-14. RCA-SATCOM.

SUMMARY

The demand for DOD and commercial satellite systems to meet many diverse communications requirements is increasing steadily. Satellite systems offer a number of distinct advantages that the military can exploit for military uses. For example, they can provide very-long-distance communication to areas that are virtually isolated or lacking in significant commercial facilities. They are capable of supporting very-high-data rates for applications involving intelligence data. For contingency operations, the use of transportable, or mobile, terminals permits rapid extension of

Figure 7-15. MARISAT.

major communications networks into new areas, or restoration of communications to locations with destroyed or damaged facilities.

Similarly, commercial or national satellite systems can provide economical and efficient expansion of existing commercial communications services. For nations with formidable terrestrial or water barriers to cable or microwave systems, communications satellites offer a practical way to quickly acquire a nationwide telecommunication system.

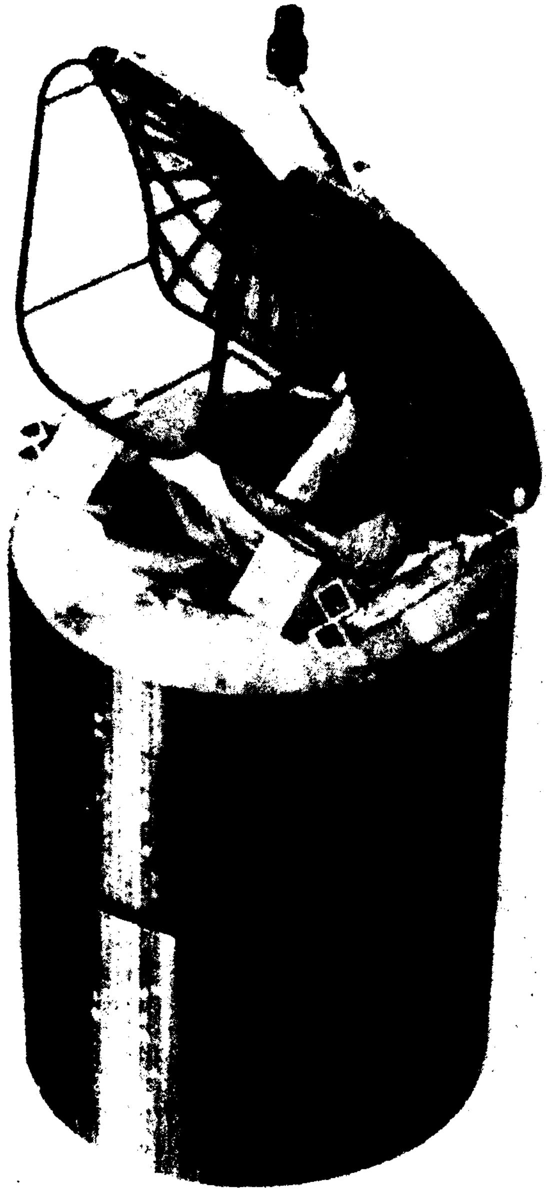


Figure 7-16. COMSTAR.

BIBLIOGRAPHY

Bond, Frederick E. *Overview—DOD SATCOM Systems.* Washington D.C.: Defense Communications Agency, US Government Printing Office, 1974.

"Communication Satellites." *Space World* (December 70):24-29.

Corliss, William R. *Putting Satellites to Work.* Washington D.C.: NASA, 1968.

Darley, R. "Orbiting Relays: Synchronous Communications Satellites." *Electronics World* 79 (February):40-41.

Dunlap, Orrin E., Jr. *Communications in Space,* New York: Harper and Row, 1970.

Filipowsky, Richard F. and Eugene I. Muehldorf. *Space Communications Systems.* Englewood Cliffs, N.J.: Prentice-Hall, 1965.

Fink, D.E. "NATO SATCOM Nears Operational Status." *Aviation Week and Space Technology* 92 (1 June 70):20-21.

"INTELSAT IV." *Space World* (April 71):59.

"INTELSAT IV—Newest Series of Commercial Communications Satellites." *Signal* 25 (April 71):43-45.

Jaffe, Leonard. *Communications in Space.* New York: Holt, Rinehart, and Winston, 1966.

Johnsen, K. "Major Decisions Face SATCOMs." *Aviation Week and Space Technology* 94 (8 March 71):53-54.

Klass, P.J. "New DOD SATCOMS Nearing Launch." *Aviation Week and Space Technology* 94 (11 January 71):40-41.

"LES (Lincoln Experimental Satellites)—An Advance in Space Communications." *Signal* 23 (February 69):18-21.

Lowry, R.M., Jr. "Military Satellite Communications." *Signal* 25 (February 69):18-21.

Mueller, George Edwin and E.R. Spangler. *Communication Satellites.* New York: John Wiley and Sons, Inc., 1964.

"Project Diana—First Man-Made Signals Bounced Off Moon." *Signal* 25 (March 71):66.

"Spacecraft Details: COMSAT Corporation." *Space World* 59 (November 68):19.

"Special Report on Communications Satellites." *Aviation Week and Space Technology* 95 (23 August 71):28-36.

"The Global Satellite Communications System." *Interavia* 26 (October 71):1165-1168.

US Congress. House. Committee on Government Operations. *Government use of Satellite Communications—1968.* 90th Cong., 2d sess., 1968. H. Rept. 1836.

Von Braun, Wehrner. "Now at Your Service: The World's Most Talkative Satellite." *Popular Science* 198 (May 71):56-57.

Wall, V.W. "Military Communication Satellites." *Astronautics and Aeronautics* 6 (4 April 68):52-57.

Chapter 8

ATMOSPHERIC PENETRATION

Not long ago, no one knew whether it was possible to bring a vehicle safely back from space. Today successful atmospheric penetration is routine. The state of the art in ballistic missile reentry vehicles has progressed rapidly and has reached the third generation of aerodynamic shapes. Scientists have developed and tested manned and unmanned lifting body craft. Because the lifting body shape permits a high degree of maneuverability during reentry, its application to ballistic missile reentry vehicles and operational spacecraft will provide considerable flexibility in mission planning.

A thorough knowledge of the factors involved in atmospheric penetration is the basis for the rapid progress in that field. This chapter will try to give the student an appreciation of these factors. The chapter includes a discussion of the characteristics of ballistic trajectories in the atmosphere, the nature of aerodynamic and heating loads, and some characteristics of lifting vehicles.

This chapter includes parameters that are not representative of modern intercontinental ballistic missile (ICBM) cones. These cones are highly streamlined and spend a short time in the atmosphere before impact. The emphasis is on vehicles that are characteristic of manned operations entering from near-earth orbits at velocities of approximately 25,000 ft/sec. When we make reference to streamlined ICBM cones for comparison, you should understand these cones are relatively blunt compared to current designs. The intent is to present fundamentals in a simplified manner.

In the vacuum of space, the thrust, velocity vector, and gravitational force govern the trajectory of a vehicle. Aerodynamic forces that cause heating due to air friction modify the vehicle in the atmosphere. Thus, as vehicles return from space flight, aerodynamic forces come into effect. We cannot use the relatively simple orbital relationships alone for predicting a trajectory. The equations of motion must consider the vehicle characteristics and nonlinear drag terms. These include atmospheric drag, a function of velocity squared (v^2) and atmospheric density (ρ), a function of altitude.

BALLISTIC TRAJECTORIES

Assume that a manned vehicle has been in an elliptical orbit about the earth, and that the vehicle has applied thrust to change its orbit to intersect the earth's atmosphere. As the vehicle enters the atmosphere, aerodynamic drag affects the trajectory. This raises some important questions for those who must recover the space vehicle and for the crew inside. The basic consideration of this section concerns the nature of the characteristics of ballistic trajectories in the atmosphere.

For an accurate analysis of a trajectory, we must consider many factors. These include vehicle characteristics, atmospheric entry angle, atmospheric density, variation of density with altitude, wind, earth rotation, earth curvature, and gravity. Since some of the factors such as velocity and density are changing constantly during the ballistic flight in the atmosphere, we consider the analysis in very small increments. Such computations are laborious when done manually. Therefore, we use computers to analyze trajectories. Even the computer does not produce exact solutions because we never know the predominant factors, velocity and density, exactly. The purpose of this section is to analyze the forces that influence motion of a vehicle.

Geometry and Assumptions

The three forces acting on ballistic vehicles as they penetrate the earth's atmosphere are drag, weight, and centrifugal force. Figure 8-1 shows these forces and the geometry of the trajectory.

We consider the vehicle to be in ballistic flight through the earth's atmosphere when it is as shown in figure 8-1. The vehicle has a velocity (v) at a distance (r) from the center of the earth. Measuring the angle that the trajectory makes with the plane of the local horizon determines the direction of the vehicle as it enters the atmosphere. We call this the entry angle (ϕ). The radius of curvature (r_c) is perpendicular to the velocity vector at the point of consideration. The drag force (D) acts opposite to and along the velocity vector. The weight of the vehicle (W) acts toward the center of the earth. Consequently, we define ballistic reentry as entry in which the only effective external forces acting on the vehicle are aerodynamic drag and gravity. Throughout the remainder of this chapter, we will assume that the earth is not rotating; there is no wind; and the gravitational acceleration, g, is constant at 32.2 ft/sec^2.

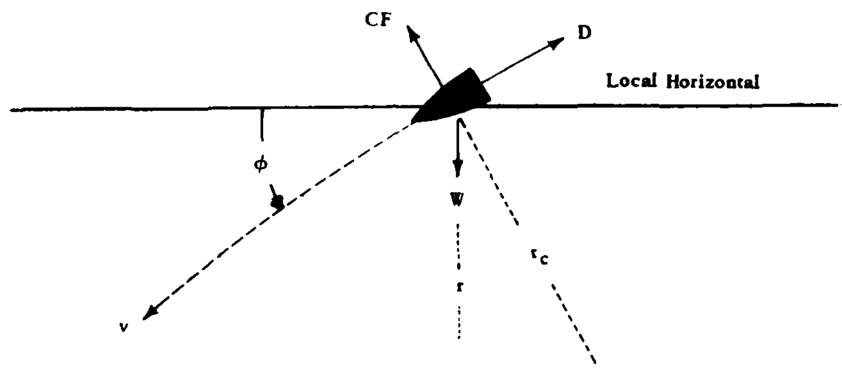


Figure 8-1. Geometry of a ballistic trajectory.

Equations of Motion (Ballistic Trajectory)

To analyze the motion of a vehicle in flight through the earth's atmosphere, start with Newton's Second Law of Motion. This law states that the unbalanced force on a body is equal to the mass times the acceleration of the body expressed as:

$$F = Ma$$

The force, *F*, is in lb; *M* is in lb-sec^2/ft or slugs ($M = \frac{W}{g}$,

where *W* is weight in lb and *g* is the acceleration of gravity in ft/sec^2), and *a* is the acceleration of the body in ft/sec^2. The change in velocity, Δv, divided by the corresponding increment of time, Δt, is equal to the acceleration:

$$a = \frac{\Delta v}{\Delta t}$$

During ballistic flight, the force, F, along the trajectory is an algebraic sum of two forces, the aerodynamic drag force (D) and a component of gravity (W sin ϕ).

The drag force acts opposite the direction of flight. We can compute the drag force from the basic aeronautical equation:

$$D = \frac{\rho v^2 C_D A}{2}$$

Where:

ρ = the density of the atmosphere and is expressed in slugs/ft^3. At sea level ρ is 0.00237 slugs/ft^3.

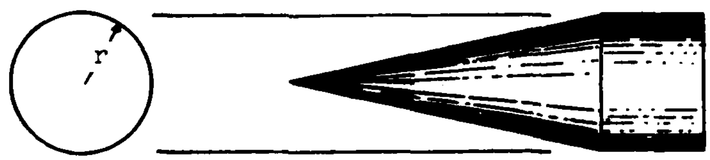
(NOTE: The units of a slug are $\frac{\text{lb-sec}^2}{\text{ft}}$ from $\frac{W}{g}$, so that $\rho = \frac{W}{(g)(ft^3)} = \frac{lb - sec^2}{ft^4}$)

v^2 = the velocity squared and the units are $\frac{ft^2}{sec^2}$.

C_D = the coefficient of drag and is a dimensionless number that reflects the shape of the vehicle. The more streamlined (less air resistant) vehicles have smaller C_D values than less streamlined vehicles. Actual values depend on many factors and vary with specific conditions. Typical values for the more streamlined shapes would be .5 and for the blunt shapes .8.

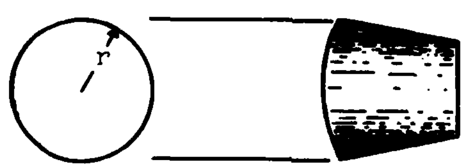
A = the effective frontal area of the vehicle exposed to the air stream and the units are ft^2. For example, the area of a sharp-nosed conical vehicle is the projected area:

$A = \pi r^2$

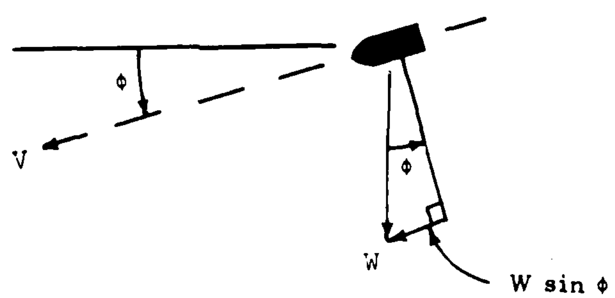


The effective frontal area is expressed in the same way.

$A = \pi r^2$



The component of gravity along the flight path is in the direction of flight and, referring to figure 8-1, we can express it as



W sin ϕ equals the component of gravity acting along the flight path, where W is the weight of the vehicle in pounds. Substitute these equivalents into Newton's Second Law, F = Ma, and develop an equation as follows:

$$F = Ma$$

$$F = -D + \text{gravity component} = -D + W \sin \phi$$

$$M = \frac{W}{g}$$

$$a = \frac{\Delta v}{\Delta t} \text{(valid for a small segment of the trajectory)}$$

$$D = \frac{\rho v^2 C_D A}{2}$$

Now:

$$F = -\frac{\rho v^2 C_D A}{2} + W \sin \phi = \frac{W}{g}\frac{\Delta v}{\Delta t}$$

Solving for $\frac{\Delta v}{\Delta t}$:

$$\frac{\Delta v}{\Delta t} = -\frac{g \rho v^2 C_D A}{2W} + \frac{g W \sin \phi}{W}$$

Simplifying:

$$\frac{\Delta v}{\Delta t} = g \left[-\frac{\rho v^2}{2\left(\frac{W}{C_D A}\right)} + \sin \phi \right] \tag{1}$$

The above equation determines the change in velocity along a small segment of trajectory. It indicates that for a small time increment, Δt, there is a corresponding change in the velocity of the vehicle. The magnitude of the Δv is a function of all the other values shown. Some of these other values (ρ, sin ϕ, and v) are constantly changing during the flight.

A very important parameter appearing for the first time in equation 1 is the ballistic coefficient, $\frac{W}{C_D A}$. This is a common parameter that appears in all studies of atmospheric penetration and plays a very important role in the behavior of a vehicle. It is simply the ratio of vehicle weight to a factor describing the degree of streamlining (C_D) and the effective frontal area of the vehicle as defined above. For example, a ballistic reentry vehicle (R/V) has a very large ballistic coefficient of 1,000 pounds per square foot or more when compared to that of the blunt-nosed Apollo capsule with a ballistic coefficient of approximately 100 pounds per square foot.

To appreciate the complexity of the computation, examine a small segment of a trajectory (see fig. 8-2). Assume that we desire the value of the change in the velocity between points one and two in a trajectory. In the small segment between point one and two in figure 8-2, ρ and ϕ do not change significantly. We assume the trajectory to be a straight line. Thus, we can write equation 1 as

$$\Delta v = \Delta t\, g \left[-\frac{\rho_1 v_1^2}{2\left(\frac{W}{C_D A}\right)} + \sin \phi_1 \right] \tag{2}$$

We can use equation 2 to analyze the velocity change when we know the conditions at point one. An example should clarify the equation and its use. Consider a manned vehicle entering the earth's atmosphere under the following conditions.

Given:

$\Delta t = 1$ second

$g = 32.2 \text{ ft/sec}^2$

altitude = 480,000 ft (80 NM)

$\rho_1 = (4.12)(10^{-12})$ slugs/ft^3*

$v_1 = 23{,}000$ ft/sec

$\dfrac{W}{C_D A} = 100 \text{ lb/ft}^2$

$\phi_1 = 5$ degrees ($\sin \phi = .087$)

Solve for: Δv, where $\Delta v = v_2 - v_1$

Substituting into equation 2,

$$\Delta v = (1)(32.2)\left[-\frac{(4.12)(10^{-12})(23{,}000)^2}{(2)(100)} + .087\right]$$

$$= 32.2\,[\,-.0000109 + .087] \approx 32.2\,[.087]$$

$$= +2.7 \text{ ft/sec}$$

$$v_2 = v_1 + \Delta v = 23{,}000 + 2.7 = 23002.7 \text{ ft/sec}$$

These calculations demonstrate that for this particular vehicle ($\frac{W}{C_D A} = 100$) and the low atmospheric density, the magnitude of the drag force is insignificant at high altitudes (80 nautical miles). In fact, the drag force is so low that the velocity is increasing by a very small amount due to the small gravity component.

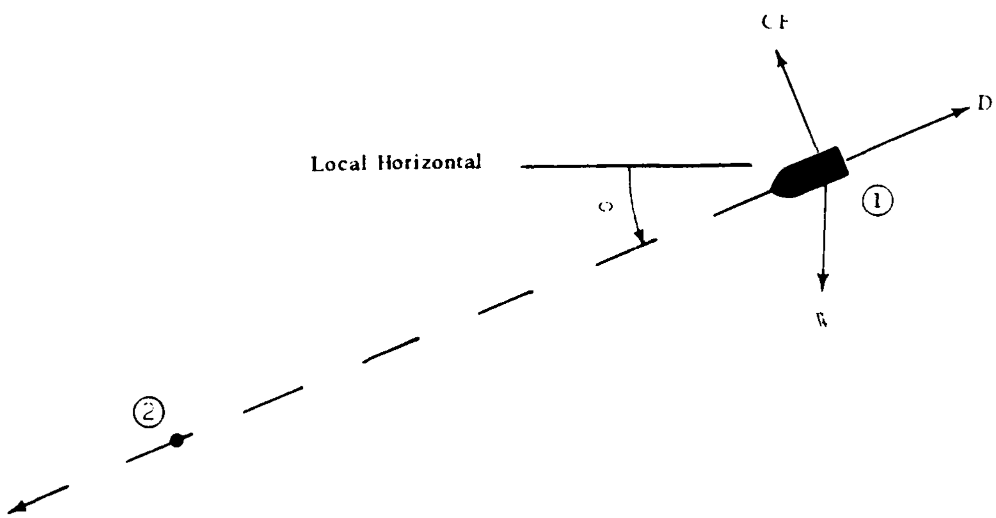


Figure 8-2. A segment of a trajectory.

*ARDC Model Atmosphere

HEATING AND DECELERATION

Heating rates and deceleration loads are the most important effects of entry into the atmosphere. These effects are most severe when there is a combination of high atmospheric density and high vehicular velocity. The most critical condition for a vehicle upon entry would involve steep entry angle and high approach velocity.

A vehicle approaching the earth's atmosphere possesses a great amount of potential energy (PE) and kinetic energy (KE). While penetrating the earth's atmosphere, the vehicle loses very little PE in relation to the loss in KE. Therefore, scientists consider all lost energy as a KE loss. KE is a function of v^2. Thus, due to the high entry velocities, the vehicle must dispel large amounts of KE as it reduces velocity to a safe impact value for manned vehicles. It is vital to dispel the energy to prevent vehicle or payload destruction.

The velocity of the vehicle at atmospheric entry depends on the mission. If the spacecraft initiates reentry from near-earth orbit, then the velocity at atmospheric entry closely approximates that of the orbital velocity. The velocity of a reentering ICBM warhead is approximately the same as the velocity at burnout of the booster. Entry from lunar missions involves velocities of 35,000 to 36,000 ft/sec. Regardless of the mission, the entry velocity will be very high and only very large retrothrust forces will change it a significant amount. This would lead to severe payload penalties. However, the spacecraft could use small retrothrust forces that allow a shallow entry angle at a lesser cost in terms of weight. Instead of large retrothrust forces, the vehicle uses the atmosphere to slow down. A shallow entry angle tends to limit to high altitudes the region of high velocity. However, the entry angle influences other factors. Shallow entry angles reduce atmospheric peak heating rates and vehicle peak deceleration loads. Vehicle heat loads may be greater for shallow entry angles and impact accuracy is less with shallow entry angles.

The heating and deceleration loads are a function of entry angle, velocity, and density. The remainder of this section will deal with the interrelationship among these factors and their influence upon heating and deceleration.

Heating

Assume a manned vehicle enters the earth's atmosphere at 24,000 ft/sec and uses atmospheric braking. Even under optimized configuration and entry angle, approximately 3,000°F will be the minimum vehicle surface temperature that we can expect during peak heating. The temperature could reach 20,000°F for steep entry angles. Two questions that arise are: what heats the vehicle and how is it protected from the intense heat?

During atmospheric penetration, aerodynamic drag transforms KE into thermal energy, heating the air surrounding the vehicle. Simply, the penetrating vehicle gets hot because it is close to the heat source, namely, the air. The amount of heat transferred to the vehicle depends on the characteristics of the air flow near it. If all the heat in the air were transferred into the vehicle, there would be more than enough to vaporize the vehicle unless the designers constructed the vehicle from a material that could either withstand extremely high temperatures or could eliminate the heat as it is produced. The following paragraphs will give a method of determining the approximate heating without going into a complex computer-derived solution.

Atmospheric heating rate. Theory holds that the heating rate is a function of the change in KE of the vehicle. We can evaluate the rate of KE change by using the following equation for KE:

$$KE = \frac{Mv^2}{2}$$

In a small segment of the trajectory, the change in KE (ΔKE) per unit of time is:

$$\frac{\Delta KE}{\Delta t} = \frac{KE_2 - KE_1}{\Delta t}$$

$$KE_1 = \frac{Mv_1^2}{2}$$

$$KE_2 = \frac{Mv_2^2}{2} = \frac{M(v_1 + \Delta v)^2}{2} = \frac{M(v_1^2 + 2v_1\Delta v + \Delta v^2)}{2}$$

since $\Delta v^2 << v_1$

$$KE_2 = \frac{M(v_1^2 + 2v_1\Delta v)}{2}$$

$$\therefore \frac{\Delta KE}{\Delta t} = \frac{Mv_1^2 + 2Mv_1\Delta v - Mv_1^2}{2\Delta t} = \frac{Mv_1\Delta v}{\Delta t}$$

Now, we will relate the above equation to an expression containing the aerodynamic drag term by use of equation 1.

$$\frac{\Delta v}{\Delta t} = g\left[-\frac{\rho v^2}{2}\left(\frac{C_DA}{W}\right) + \sin\phi\right]$$

In the area of high-energy conversion, where Δv is high, $\sin\phi$ is very small in relation to other parameters and we can neglect it. This simplifies the previous equation to

$$\frac{\Delta v}{\Delta t} = -\frac{\rho v^2}{2} g\left(\frac{C_DA}{W}\right).$$

Multiply both sides of the above equation by Mv and obtain:

$$\frac{Mv\Delta v}{\Delta t} = -\frac{\rho v^3}{2} g\left(\frac{C_DA}{W}\right) M = -\frac{\rho v^3}{2}(C_DA)$$

Now equating the two terms that equal $\frac{Mv\Delta v}{\Delta t}$:

$$\frac{\Delta KE}{\Delta t} = -\frac{\rho v^3}{2}(C_DA). \qquad (3)$$

Surface heating rate. Equation 3 is an expression that shows the heating of the air, as transformed from KE, at a point in the trajectory. Thus, for a given configuration (a known C_DA), and known velocity and altitude (density) history during a penetration, we can use the equation to determine the heating trend and the point (altitude and velocity) of maximum heating. A plot of these values (versus time or altitude) will depict the heating trend and reveal the area of peak heating. We make no attempt to compute the actual heating rates or temperatures. This text is concerned with the predominant heating factors.

Equation 3 shows the thermal energy transferred to the air. The interest lies in knowing how much heat transference to the vehicle occurs. It is important that the air, not the vehicle, absorb as much of this heat as practicable. The energy conversion factor (ECF) is the fraction of converted kinetic energy that enters the vehicle as heat. The heat input determines the type of vehicle protection required and the corresponding payload weight penalty imposed. This fraction of converted energy is

$$\frac{\text{Heat absorbed by the vehicle}}{\text{Total heat generated}} = ECF.$$

The magnitude of the ECF depends primarily on the vehicle shape, velocity, and atmospheric density (altitude). Figure 8-3 shows the effect of shape and density on the ECF and the type of airflow for typical ballistic trajectories. This figure indicates that molecules of air impinging directly on the surface and transferring heat to the vehicle would transfer as much as one-half of the total heat generated. Fortunately, at these extremely high altitudes, the atmospheric density is very low. Thus, the KE transfer is correspondingly low. Large ECFs may occur at lower altitudes for slender shapes.

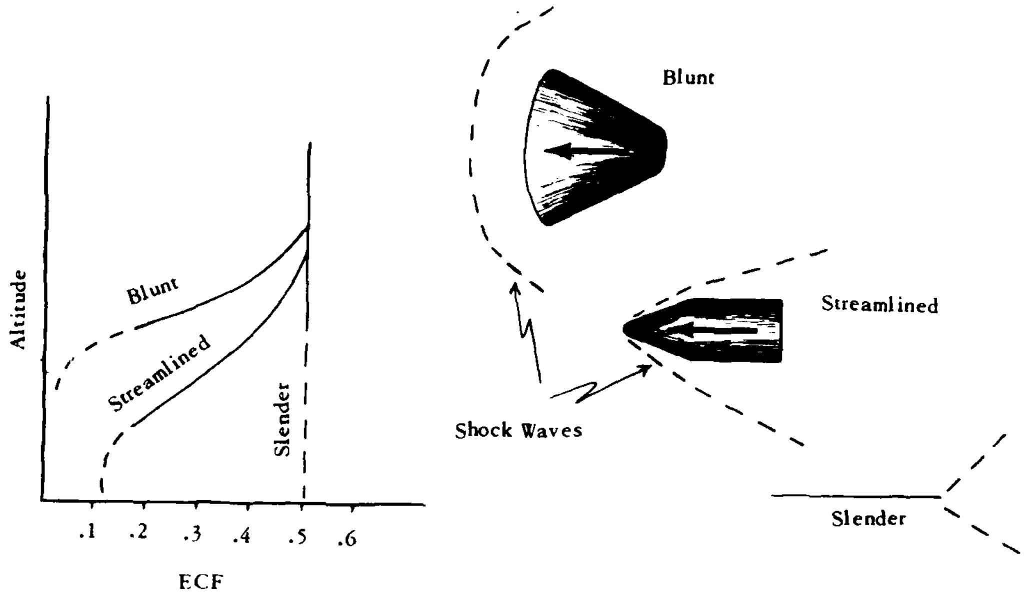


Figure 8-3. Parameters affecting energy conversion factor.

Figure 8-4 shows details that reveal why the shape of the vehicle has such strong influence on the ECF. This figure shows the air flow pattern around a blunt (high-drag) vehicle and a streamlined (low-drag) vehicle.

With the blunt vehicle, there is a detached shock wave ahead of the vehicle. It is nearly normal (perpendicular) to the velocity vector. The air between the surface and the shock wave is hot and moves slowly with respect to the vehicle. The velocity of the air in the boundary layer (a thin layer of air next to the vehicle) is very low. The boundary layer acts as insulation. However, some heat does transfer from the hot air, through the boundary layer to the surface of the vehicle.

With a *streamlined* vehicle, the shock wave is almost parallel to the air stream, and the air in the boundary layer is moving relatively fast along the surface of the vehicle. There is a large change in velocity (velocity shear) between the vehicle surface and outer edge of the boundary. *The viscosity of the air and the shearing action* cause the maximum temperature to occur within the boundary layer.

The rate of heat transfer is different for vehicles of different configuration. We can see that the temperatures created by the streamlined vehicle are high near the surface. Thus, the streamlined vehicle receives more heat transference than does the blunt vehicle. This is the reason the ECF is higher for the slender vehicle.

The character of the air flow in the boundary layers influences the ECF. It does not matter whether the air flow is laminar or turbulent. In laminar flow, the air moves smoothly in layers (lamina). Turbulent flow has the characteristic of irregular, eddying, or fluctuating flow. The turbulent boundary layer occurs in the lower atmosphere and allows a much higher rate of heat transfer to the vehicle.

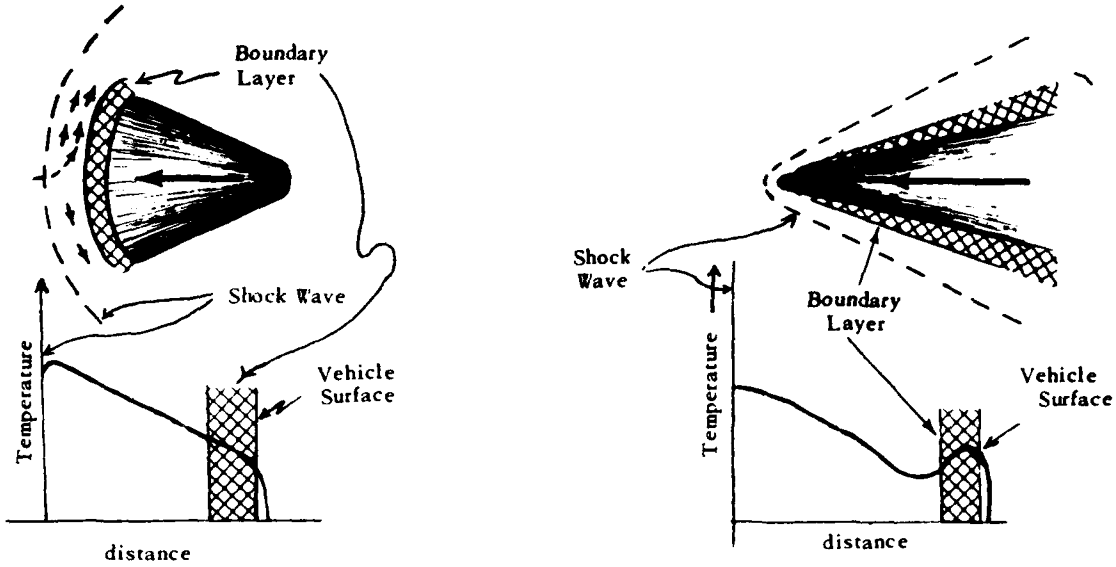


Figure 8-4. Aerodynamic heating.

Figure 8-5 shows a typical variation of the ECF for a blunt vehicle penetrating the earth's atmosphere. The ECF varies from about 0.5 at high altitudes to about 0.01 at 100,000 feet. At lower altitudes, the ECF increases again due to a turbulent boundary layer condition, as the dashed portion of the curve indicates.

Protection from heat. Scientists and engineers have proposed a great many techniques for protecting the vehicle from the intense heat during the atmospheric penetration. All these techniques fall into one of two general classes, namely, radiation cooling and heat absorption systems.

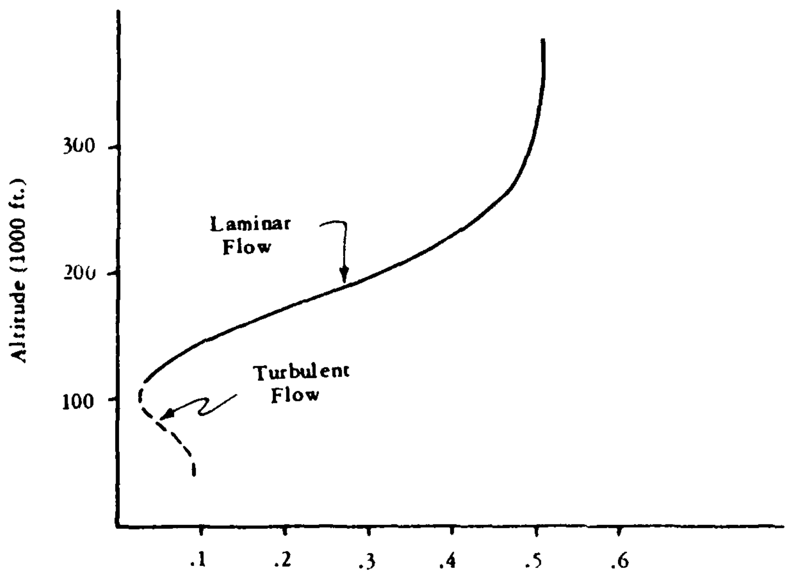


Figure 8-5. ECF versus Altitude.

If the surface heating rate is low, radiation cooling may occur. The main characteristic of this method is the radiation of some of the heat back into the atmosphere. The maximum temperature that the vehicle can withstand is the limiting factor in this type of cooling. In this process there is a point where the amount of heat absorbed by the vehicle equals the amount radiated back into the atmosphere. For a given reentry condition and altitude an equilibrium temperature then exists, and the temperature of the structure has reached a maximum. Radiation cooling was the method used for protecting the lifting surfaces of the Asset glider. Such vehicles decelerate slowly and have high velocities for a long period of time at high altitude. Consequently, they undergo low-heat rates but high-heat loads.

A heat sink system uses a large mass of metal (usually copper) to absorb a great quantity of heat before melting. Scientists used this technique early in the ballistic missile reentry vehicle program, but it proved too heavy for today's operational vehicles.

Transpiration implies the boiling off of a liquid. For example, water boils at about 70°F at the low pressures encountered in space. Therefore, it would maintain the part of the vehicle that it was protecting at 70°F. However, as the pressure increased during reentry, the boiling temperature would increase as well.

Ablation is the wearing away, melting, charring, or vaporizing of the surface material. The vaporizing process leaves a thin vapor layer near the vehicle, thereby insulating the vehicle from the intense heat. The wearing away and melting process carries some heat away from the vehicle. Examples of ablative materials are reinforced plastics, resin, fiberglass, and cork.

Deceleration

A spacecraft returning from a low-earth orbit enters the earth's atmosphere at 24,000 ft/sec. The spacecraft must reduce its velocity greatly prior to landing. The reduction of velocity per unit of time is the rate of velocity change, or acceleration. If the velocity is decreasing, the acceleration has a negative value and we call it deceleration. This is the case when a ballistic vehicle penetrates the atmosphere. The larger the change in velocity per unit time, the greater the deceleration. Therefore, consider the very high change in velocity and corresponding deceleration loads that may range up to hundreds of *G*s with the peak *G* determined by the entry conditions. Knowing that each system has a maximum allowable *G*, we face the question of what factors influence the peak deceleration loads during penetration.

The motion of a vehicle during atmospheric penetration depends on vehicle velocity, atmosphere entry angle, ballistic coefficient, lift characteristics, and atmosphere density. During atmospheric entry the vehicle may pitch, yaw, roll, and experience accelerations in the longitudinal and lateral directions. Because it is of greater magnitude, the longitudinal deceleration of the vehicle is the only dynamic motion considered in this text.

The most important question of vehicle dynamics concerns the magnitude of maximum deceleration compared with the earth's gravitational acceleration, *G*. To determine the answer, use equation one to plot deceleration history.

$$\frac{\Delta v}{\Delta t} = g\left[-\frac{\rho v^2}{2\left(\frac{W}{C_D A}\right)} + \sin\phi\right]$$

For lifting vehicles, we can obtain the value of C_D from a curve of C_D versus C_L. The maximum deceleration for a lifting vehicle depends on the entry conditions and the lift-drag ratio (L/D). The velocity decrease extends over a longer period of time than for a ballistic vehicle, with consequently lower deceleration. Thus, both a shallow entry angle and aerodynamic lift serve to decrease the deceleration during atmospheric penetration. Figure 8-6 shows the maximum longitudinal deceleration as a function of initial entry angle and L/D. Note that relatively small L/D values from vehicles developing lift can obtain a large reduction in maximum deceleration. The longitudinal

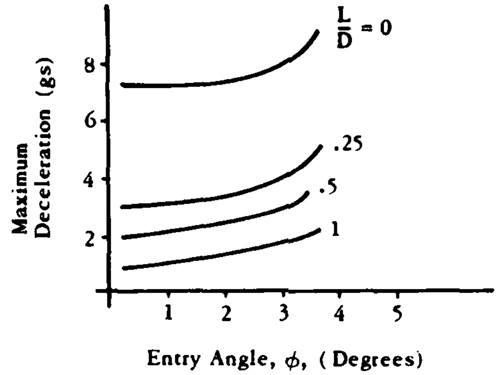


Figure 8-6. Maximum deceleration versus entry angle for various values of $\frac{L}{D}$.

deceleration of a lifting vehicle is very low and is not a problem. Therefore, the remainder of this section concerns ballistic (nonlifting) vehicles.

We can compute easily the magnitude of deceleration. Deceleration is $\frac{\Delta v}{\Delta t}$ in equation 1. The units are in ft/sec^2. Dividing $\frac{\Delta v}{\Delta t}$ by G converts magnitude of deceleration into terms of gravitational acceleration. Since we have established the basis for computing deceleration, we will return to the example problem and compute the deceleration for the manned ballistic vehicle entering the earth's atmosphere.

$\Delta t = 1$ second was chosen, and $\Delta v = 2.7$ ft/sec was computed. $\frac{\Delta v}{\Delta t} = \frac{2.7}{1} = 2.7$ ft/sec^2

(positive value; thus, vehicle is accelerating).

In terms of the gravitational acceleration:

$$\frac{\Delta v}{g \Delta t} = \frac{2.7}{(32.2)(1)} = .08g$$

Consider the same manned vehicle at a 100,000-foot altitude travelling 8,000 ft/sec at an entry angle of 10 degrees under the following conditions:

$g = 32.2$ ft/sec^2

$\rho = (3.2)(10^{-5})$ slugs/ft^3

$v = 8{,}000$ ft/sec

$\frac{W}{C_D A} = 100$ lb/ft^2

$\phi = 10$ degrees ($\sin \phi = .174$)

$$\frac{\Delta v}{\Delta t} = g\left[-\frac{\rho v^2}{2\left(\frac{W}{C_D A}\right)} + \sin\phi\right] = g\left[-\frac{(3.2)(10^{-5})(8000^2)}{(2)(100)} + .174\right]$$

$$\frac{\Delta v}{\Delta t} = g(-10.2 + .174)$$

$$\frac{\Delta v}{\Delta t} = -10.026\, g \text{ (negative value } \therefore \text{ vehicle is decelerating)}$$

By analyzing the above, it is evident that the direct contribution of sin ϕ is negligible in the area of high deceleration. Even at a 90 degree entry angle the maximum acceleration due to gravity would be 1 g (sin $\phi = 1$). However, the entry angle shows its real influence indirectly. An examination of equation 1 indicates that the entry angle is the only variable in the equation that we can control realistically. The other factors (gravity, velocity magnitude, atmospheric density, and the ballistic coefficient) are all fixed, and we can change only the entry angle for a specific mission.

Theoretically, the maximum deceleration is fixed for any fixed entry angle and velocity magnitude. The only thing that we can change is the altitude at which the maximum occurs. We can change that altitude only by changing the ballistic coefficient, $\frac{W}{C_D A}$. Vehicles with large ballistic coefficients merely penetrate further into the atmosphere before experiencing maximum deceleration. Regardless of difference in shape, size, and weight, two different vehicles approaching the atmosphere at the same velocity magnitude and angle will experience the same maximum deceleration. Only their trajectories and altitude of maximum deceleration will be different. Figure 8-7 shows the influence of changing the ballistic coefficient for a shallow entry angle (in the order of 1 degree). Note that both vehicles experience the same magnitude of deceleration. The basic difference is that the vehicle with the smaller ballistic coefficient encounters maximum deceleration at a higher altitude.

Theoretically, a ballistic reentry vehicle entering the earth's atmosphere with a given velocity magnitude has a peak deceleration that only the entry angle determines. If we increase the entry angle, the magnitude of maximum deceleration increases. In other words, the steeper the angle, the higher the maximum deceleration will be. Figure 8-8 shows the effect of entry angle on the maximum deceleration for a vehicle such as the Project Mercury capsule. The shallow entry angles cause the deceleration to start higher in the atmosphere and last over a long period of time. This results in a smaller peak deceleration. If we used a vehicle with a smaller ballistic coefficient, these curves would shift to a higher altitude with identical peak decelerations for the corresponding entry angles. Conversely, a vehicle with a larger ballistic coefficient would cause the curves to shift downward.

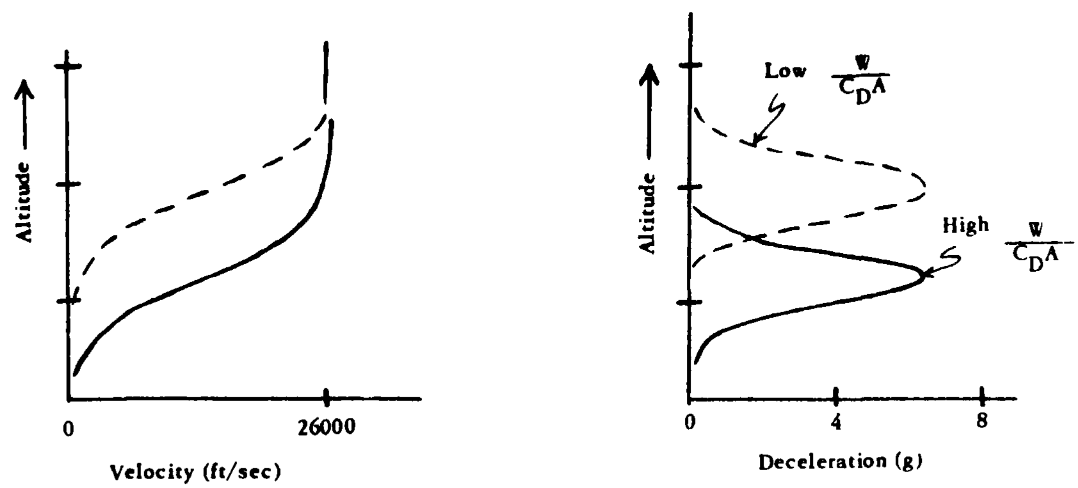
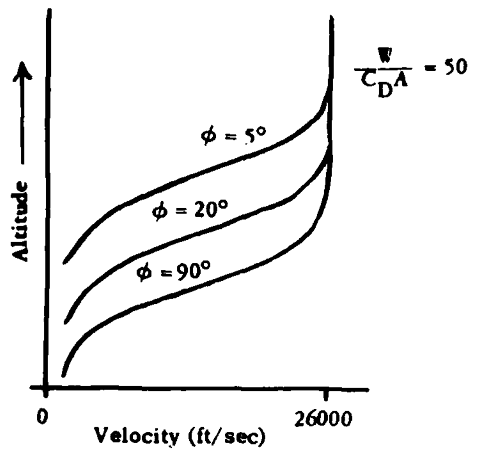


Figure 8-7. Influence of ballistic coefficient for given entry angle.



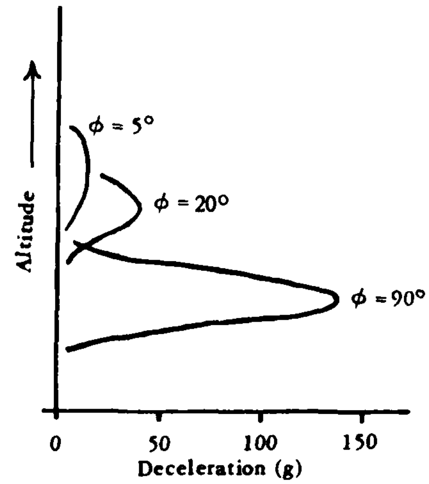


Figure 8-8. Influence of entry angle on velocity and deceleration.

LIFTING VEHICLES

We have centered the discussion thus far mainly on ballistic vehicles. When viewed from a military standpoint, these vehicles have an inherent disadvantage. This is our inability to control the vehicle through the atmosphere to arrive at an exact impact point. For example, after use of retrothrust to deorbit a vehicle, there is little that we can do to control the final destination point of the vehicle. Its impact point depends solely on ballistic parameters.

Advantages and Disadvantages

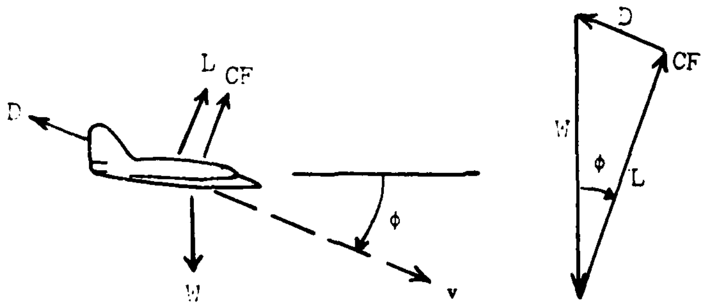
The lifting vehicle allows a more gradual descent through the atmosphere. This lifting capability leads to several potential advantages depending on the specific mission of the vehicles. These advantages include low deceleration loads, low heating rates and the ability to maneuver in the atmosphere. This maneuverability in the atmosphere allows the vehicle to land at a predetermined point as well as leaving a landing footprint. The vehicle requires a minimum of recovering forces. It can create a deep atmospheric reentry corridor and has the capability of synergetic plane change.

There are some disadvantages as well. The lifting surfaces add weight. This reduces the overall range and useful payload for a given booster. The longer period of deceleration may cause the total heat load of the vehicle to be much greater. There has been far more experience with ballistic than lifting vehicles. Obviously, this has left lifting technology behind ballistic technology.

These are some of the advantages and disadvantages of a lifting vehicle. We will not discuss all of these. Rather, the purpose of this section is to explore the ground range capability of a lifting vehicle after it enters the earth's atmosphere.

Lift and Area Factors

In analyzing the glide of a lifting vehicle, the four forces acting on the vehicle during penetration are weight, drag, lift, and centrifugal force (CF). Since this does not accelerate the motion, the equilibrium conditions lead to the equations $L + CF = W\cos\phi$ and $D = W\sin\phi$.



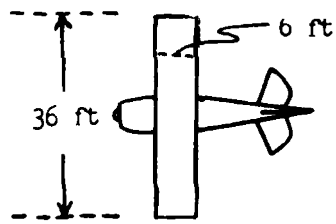
We have defined all of these terms in previous discussions except the lift (L). This is the prime consideration in the lifting vehicle. We can use a basic aeronautical equation to determine lift:

$$L = \frac{\rho v^2 C_L S}{2}$$

We have discussed all of the parameters in this equation except the life coefficient, C_L, and the area, S. The lift coefficient is a dimensionless number and reflects the lifting capability of a particular surface at a given angle of attack relative to the airstream. C_L is a function of the shape of the lifting surface. For example, we associate a very high C_L with thick wings that have a high lift for a given wing area. Usually, these are found on slow-flying aircraft. High-speed aircraft cannot use thick wings due to the excessive drag caused by the thickness. Thin wings characteristically have a very low C_L, and they normally employ flaps, slots, or slats to augment the wing lift during takeoff and landing.

The area, S, is the area of the lifting surface. For example, an airplane with rectangular-shaped wings has a wing area equal to the wing span times the width of the wing.

$$S = 6 \times 36 = 216 \text{ ft}^2$$

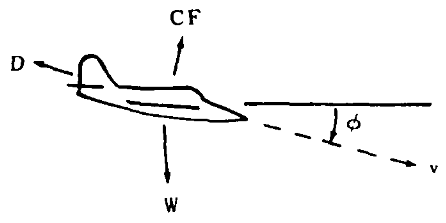


Glide Analysis

There are many analytical approaches used on the gliding flight of a vehicle. Most textbooks use theoretical mathematical approaches. However, in this text, we will use a less complex approach to provide familiarization with the factors that influence the ground range of lifting vehicles.

We will analyze the glide in three steps. The first is the ballistic period. This includes the portion of the flight down to 250,000 feet. We consider this as ballistic (nonflying) flight. In this case, we define 250,000 feet as the atmospheric entry altitude. The second is the altitude hold period. As the vehicle enters the 250,000-foot level, the pilot will maintain altitude until the airspeed diminishes to the best glide speed. The third is the descending glide period. As the vehicle approaches the best glide speed, the pilot initiates a descending glide. The pilot maintains the vehicle at the attitude that corresponds to $\left(\frac{L}{D}\right)_{max}$, which is the flight attitude that will result in a minimum glide angle and maximum horizontal distance. We do not consider the terminal maneuvers necessary to recover the vehicle part of the glide phase. Therefore, we will not discuss them.

Ballistic period. In the ballistic period, the lift vehicle enters the earth's atmosphere at near-orbital speed, and at a shallow angle. This attitude remains constant until the vehicle descends to approximately 250,000 feet. At this point the atmosphere has sufficient density to support lift at reentry velocities.



Altitude hold period. During the altitude hold period, the pilot maintains altitude by gradually increasing the angle of attack as the speed decreases to the best glide speed. The distance travelled while the vehicle slows to optimum glide speed then becomes an important factor. Distance is the average speed multiplied by the time ($v_{avg} \times t$). We will examine speed first.

Figure 8-9 shows the external forces at the point of optimum glide speed,

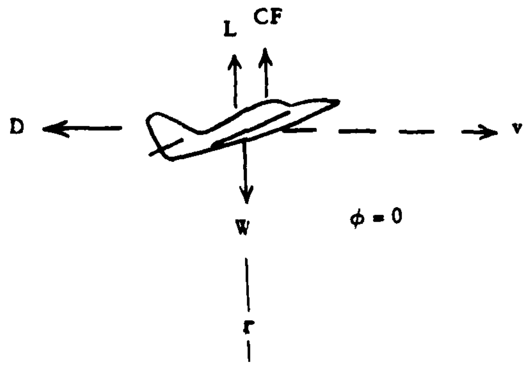


Figure 8-9. External forces at the point of optimum glide speed.

$$W = L + CF = \frac{C_L \rho v^2 S}{2} + \frac{Mv^2}{r};\ M = \frac{W}{g};\ CF = \frac{Mv^2}{r}$$

To obtain the average speed it is first necessary to solve the above relation for optimum glide speed, v_g, when we use C_L for optimum glide.

$$v_g = \sqrt{\frac{W}{\frac{C_L \rho S}{2} + \frac{W}{gr}}} \qquad (4)$$

We obtain the lift coefficient C_L from the performance curves for the particular vehicle. The average speed, v_{avg}, is the entry speed, v_e, plus optimum glide speed, v_g, divided by 2.

$$\left(v_{avg} = \frac{v_e + v_g}{2}\right)$$

The next calculation is to find the required time to slow the vehicle to optimum glide speed while maintaining altitude. To do this, use Newton's Second Law of Motion, in the horizontal direction.

$$F = Ma, \text{ from the preceding figure}$$

$$F = D = \frac{C_D \rho v^2 S}{2}$$

$$M = \frac{W}{g}$$

$$a = \frac{\Delta v}{\Delta t}$$

Substituting into the latter equation and solving for Δt (the time required to slow to optimum glide speed),

$$\Delta t = \frac{2 \Delta v W}{g \rho v^2_{avg} C_D S}; \ \Delta v = v_e - v_g$$

$$\text{Distance} = (\Delta t)(v_{avg}) = \frac{2 \Delta v \, v_{avg} W}{g \rho v^2_{avg} C_D S} = \frac{2 \Delta v W}{g \rho v^2_{avg} C_D S} \tag{5}$$

Descending glide period. In analyzing the descending glide period, we can resolve the solution for the glide angle, ϕ, by analyzing the following equations analyzed at the beginning of this section, $L + CF = W \cos \phi$ and $D = W \sin \phi$. By dividing the first of these equations into the second, we get

$$\frac{D}{L + CF} = \frac{W \sin \phi}{W \cos \phi} = \tan \phi$$

The magnitude of the centrifugal force is very small at glide speed and we can neglect CF. This leaves the solution of the glide angle in a simple form:

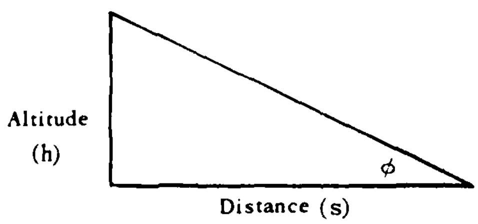
$$\tan \phi = \frac{D}{L}; \left(\text{or } \cot \phi = \frac{L}{D}\right) \tag{6}$$

This equation indicates that once we know the $\frac{L}{D}$ characteristics of a vehicle, we can approximate the optimum glide angle of the vehicle. When we know the glide angle and altitude, the distance travelled in the glide is a simple relationship of the angle:

$$\cot \phi = \frac{S}{h}$$

$$S = h \cot \phi$$

$$S = h \left(\frac{L}{D}\right) \tag{7}$$



CONCLUSION

This chapter presented some of the problems that arise during penetration of the atmosphere with emphasis on the heating and deceleration aspects. We have presented only that detail that is needed to provide some understanding of the problem and to lay a foundation for further analysis.

BIBLIOGRAPHY

Duncan, Robert C. *Dynamics of Atmospheric Entry.* New York: McGraw-Hill, 1962.

Ely, Lawrence D. *Return from Space.* Springfield, Ill.: Charles C. Thomas, 1966.

Gazley, Carl., Jr., "Deceleration and Heating of a Body Entering a Planetary Atmosphere." in *Vistas in Astronautics,* Astronautics Symposium, 1957. New York: Peragamon Press, 1958.

La Galley, Donald P. *Ballistic Missile and Space Technology Reentry and Vehicle Design.* Vol. 4, New York: Academic Press, 1960.

Loh, W.T.H. *Dynamics and Thermodynamics of Planetary Entry.* Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Morros, C.T., L.D. Ely and M.R. Smith. *Ballistic Missile and Aerospace Technology Reentry.* New York: Academic Press, 1961.

RAND Corporation. *Aerodynamic Aspects of Future Space Vehicles and Implications for Testing.* Washington D.C.: RAND Corporation, 1967.

__________. *Atmospheric Entry of Manned Vehicles.* Santa Monica, Calif.: RAND Corporation, 1960.

__________. *An Introduction to Astronautics.* Vol. 1, Santa Monica, Calif.: RAND Corporation (1958): 10.1 10.32

Truitt, Robert W. *Fundamentals of Aerodynamic Heating.* New York: Ronald Press, 1960.

Chapter 9

DIRECTED ENERGY SYSTEMS

No discussion of space today would be complete without a serious analysis of lasers and particle beams. Scientists have given these directed energy systems a great deal of attention because they have potential as long range, hard kill antisatellite and antiballistic missile weapons. To appreciate their true potential for space warfare, we must understand the physics and engineering of beam generation, propagation, and target interaction.

LASER DESIGN AND DEVELOPMENT

In less than two decades after its discovery, the laser has had a powerful impact on science and technology. The intense, coherent, monochromatic radiation produced by the laser has sharpened understanding of the physical laws describing the world and has led to achievements with radiation considered pipe dreams only a few years ago. The laser has found applications in such diverse fields as medicine and industrial welding as well as in the military arena. The success of the laser-guided bomb in Vietnam is only one example of the military uses of the laser.

Basic Ideas

Laser is an acronym for light amplification by stimulated emission of radiation. Strictly speaking, light is electromagnetic radiation in the portion of the electromagnetic spectrum visible to human eyes, roughly over a wavelength that ranges from 0.4 to 0.7 micrometers (μm). However, the phenomenon of amplification through stimulated emission is not limited to this range. Indeed, it was first observed with microwaves in wavelengths of 1.25 centimeters, and later in the infrared and X-ray portion of the spectrum. In other words, physicists and laser engineers have generalized the meaning of light to include the electromagnetic spectrum (fig. 9-1).

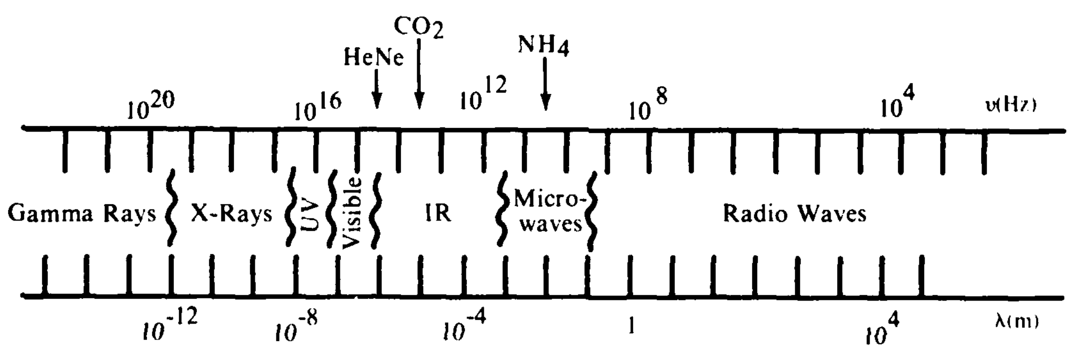


Figure 9-1. Electromagnetic spectrum.

The emission of radiation is the process by which an atom or molecule loses energy. A good example of the process is the light from an incandescent bulb. The electrical current passing through the wire filament heats the wire's molecules and transfers energy to them. Then the molecules lose energy by the emission of radiation, and give light in the process, making the bulb useful. Conversely, atoms or molecules can consume energy by the reverse process of absorption. For example, skin is not transparent. It absorbs a fraction of the sunlight that falls on it and becomes heated.

Interestingly enough, atoms and molecules can emit or absorb radiation only at certain specific wavelengths. Figure 9-2 shows some of the wavelengths of light absorbed or emitted by hydrogen. A striking feature of the process is that emission and absorption occur as a series of lines of very narrow bandwidth. One of the notable successes of modern physics is the explanation of these results in terms of the internal structure of atoms and molecules.

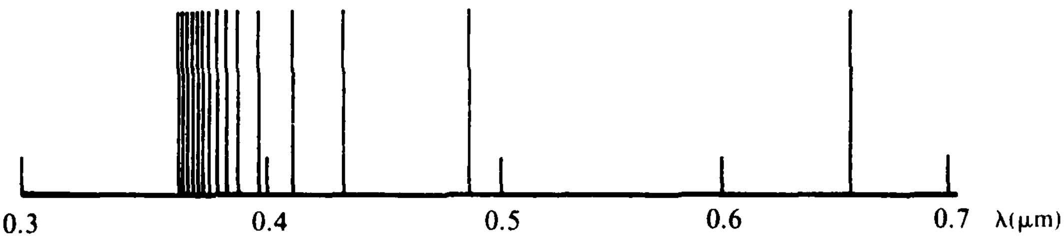


Figure 9-2. Wavelengths of light in the visible region at which Hydrogen absorbs or emits radiation.

Central to understanding the emission and absorption of light from atoms and molecules is the notion that we may consider light not only as an electromagnetic wave but also as a stream of tiny particles. This notion may seem contradictory, since waves are continuous, and particles are discrete, but we can fully explain this physical phenomena only if we accept the idea that waves behave to some extent as particles. This wave-particle duality is a key concept of modern physics.

How does light behave as a stream of particles? In its interaction with light, matter absorbs light only in discrete little "packets," which, for light of frequency ν, possesses energy $h\nu$ and momentum $2\pi\nu h/c$, where h is a constant (Planck's constant, 6.63×10^{-34} joule-sec) and c is the speed of light (3×10^8 m/sec). For example, microwaves have a frequency $\nu = 2 \times 10^{10}$ Hz, which gives energy for a little packet of microwaves $\epsilon = h\nu = 10^{-23}$ joules. That is very little energy. The average energy per molecule at room temperature is approximately 4×10^{-21} joules, or $2\frac{1}{2}$ orders of magnitude greater. Thus, a substance must absorb at least 100 packets of microwave energy per molecule to make a detectable difference in its temperature; the individual lumps are lost in the crowd. Similarly, water consists of discrete H_2O molecules, but it seems continuous simply because the molecules are so small on a detectable scale.

However, the situation is different with light at visible wavelengths. Here, $\nu = 6 \times 10^{14}$ Hz and $\epsilon = h\nu = 4 \times 10^{19}$ joules. This is two orders of magnitude greater than the average energy per molecule at room temperature. We can observe readily the discrete "kick" given to a molecule when it absorbs visible light. For instance, we see this in the photoelectric effect of light absorbed in a vacuum tube actually knocking out electrons and allowing a current to flow.

Physicists speak of the little packets of light as photons and describe frequencies in terms of radians/sec. The sinusoidal variation of light waves with ωt motivates the radian frequency $\omega = 2\pi\nu$. Usually scientists write the energy of a photon as $\epsilon = \hbar(\omega)$ where $\hbar = h/2\pi$, and its momentum as $p = \hbar k$, where $k = 2\pi/\lambda$, $\lambda = c/\nu$ is the wavelength and k is the wave number.

Since we may view light as a stream of photons of energy $\hbar\omega$ and momentum $\hbar k$, we can examine the implications with respect to its emission or absorption. In fact, the momentum of a photon is very small compared to the energy it carries, and, as a result, a photon cannot be absorbed while it

is imparting translation motion to an atom. As figure 9-3 shows, a photon might collide with a molecule of N_2 and be absorbed.

Conservation of energy requires $\hbar\omega = {}^1/_2 mv^2$, and conservation of momentum requires $\hbar k = mv$. In solving the first expression for v, we get $v = (\frac{2h\omega}{m})^{1/2} = 827$ m/sec; in solving for $\hbar\omega$, we get $\hbar\omega = 1.6 \times 10^{-20}$ joules, which is the energy of a photon from a CO_2 laser. Solving the second expression for v, we get $v = \hbar k/m = 10^{-3}$ m/sec. In other words we cannot simultaneously conserve both energy and momentum in this way. This does not mean that protons cannot be absorbed. Nature can get around this by exciting internal degrees of freedom in the molecules, so that energy can be transferred with little net velocity. For example, we might set the molecule spinning in a pattern called a rotational excitation. In this instance the individual atoms are moving, but the net momentum is zero (fig. 9-4). We might set it vibrating in a pattern called vibrational excitation (fig. 9-5), or we might excite the electrons from an inner orbit to an outer one in a pattern called an electronic excitation (fig. 9-6). In these ways, energy can be absorbed without imparting momentum, and photons can be absorbed.

There is a complicating factor in all of this: the internal motions of atoms and molecules are quantized with only discrete values. For example, figure 9-7 shows the allowed energy levels of two atoms of nitrogen (N_2) molecules. The implication is that N_2 will absorb or emit protons only if they connect two allowed energy levels. This explains the discrete absorption and emission spectra shown in figure 9-2.

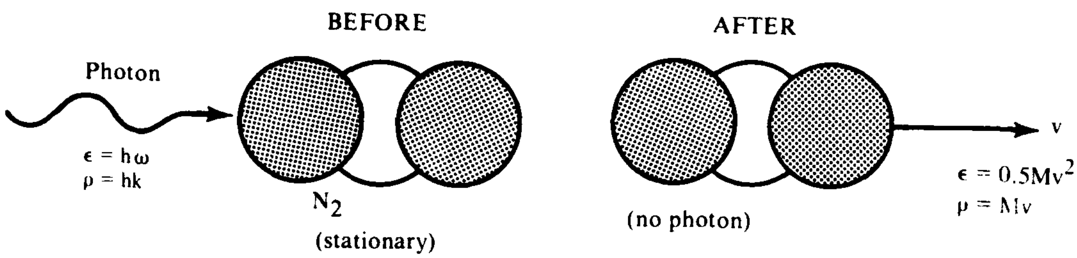


Figure 9-3. Photon absorption.

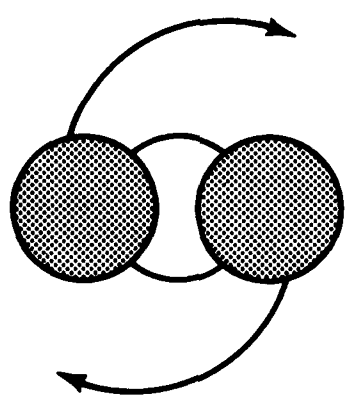


Figure 9-4. Rotational excitation.

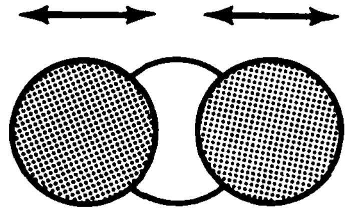


Figure 9-5. Vibrational excitation.

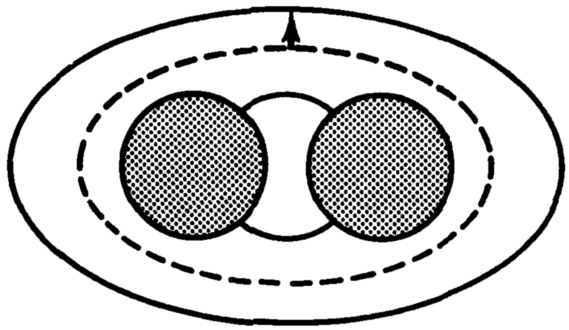


Figure 9-6. Electronic excitation.

In contrast to spontaneous emission, stimulated emission of radiation occurs when photons incite the emission of other photons. In spontaneous emission, photons are emitted without outside help and are responsible for most of the light seen in everyday experience. Matter, heated to a high temperature, randomly emits protons. An example of this would be the sun or a lamp filament. The atoms try to cool off by losing energy in the form of photons. The resulting light goes out in all directions, has a variety of wavelengths corresponding to the many possible transitions, and has no coherent spatial or temporal structure. However, in a stimulated emission, a photon encounters an atom in an excited state and stimulates it to emit a photon. This stimulation can occur only if the original and the emitted photon have the same frequency; that is, if there is a resonance between the frequency of the incident photon, ω, and the difference in energy between the excited and unexcited atom $\Delta\epsilon$, is such that $\Delta\epsilon = \hbar\omega$. If this happens, then the emitted photon will have the same frequency as the first, will be in phase with the first, and will be emitted in the same direction as the first.

An example of this might be a molecule vibrating back and forth at a frequency, somewhat like two balls on a string (fig. 9-8). The structure of allowed energies is $\epsilon_n = (n + {}^1/_2)\Delta\epsilon$, where $\Delta\epsilon$ is an energy difference established by the "stiffness" of the spring. This corresponds to vibrational frequencies $\omega_N = (N + \frac{1}{2})\Delta\epsilon/\hbar$. If some light comes in with some frequency resonant with the energy difference between two levels of the molecule, we can reasonably expect something to happen. If the molecule vibrates at high frequency, it may be induced to go to a lower frequency and give up energy in the form of a photon resonant with the first. This is stimulated emission. On the other hand, if the molecule vibrates at a low frequency, it may be induced to go to a higher frequency and absorb the first photon. Absorption and stimulated emission are competing processes; the former consumes photons, and the latter creates them. Figure 9-9 illustrates the three processes of absorption, spontaneous emission, and stimulated emission.

Light amplification is possible if the rate of stimulated emission exceeds absorption. How can we ensure this rate? It should be clear that the probability of stimulated emission should be proportional

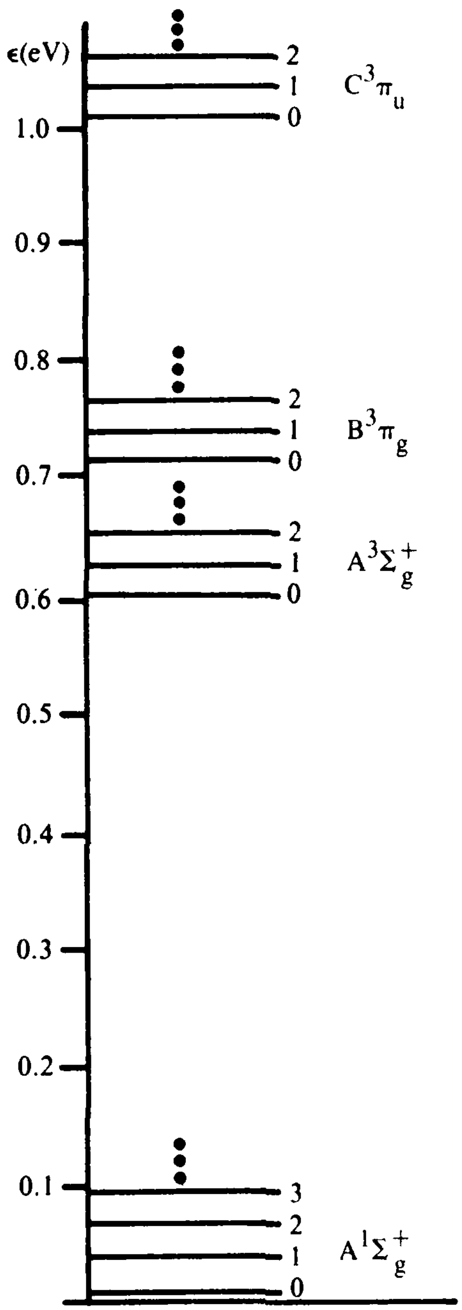


Figure 9-7. Energy level structure of N_2.

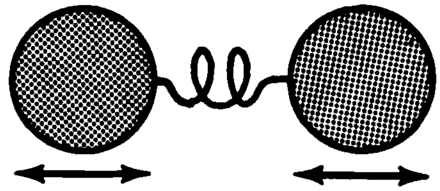


Figure 9-8. Vibrational motion.

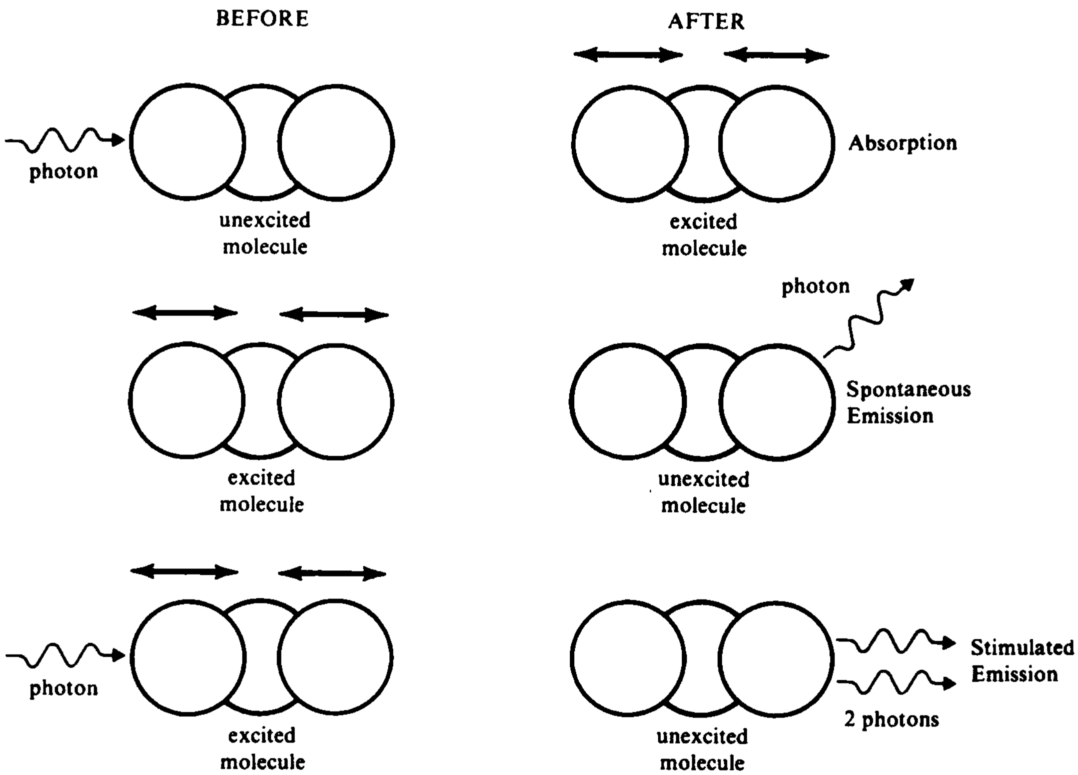


Figure 9-9. Emission and absorption.

to the number of excited atoms (N^*), times the likelihood per atom of stimulated emission (B_S), times the number of photons (n_p) around to act as stimulators. Similarly, the probability of absorption is proportional to the number of unexcited atoms (N) times the likelihood per atom of absorption (B_A) times the number of photons (n_p) available for absorption. In other words,

$$N^* B_S n_p > N B_A n_p \,. \tag{1}$$

Einstein showed in 1917 the $B_S = B_A$, therefore, we can conclude from equation 1 that light amplification requires

$$N^* > N \,. \tag{2}$$

We know the situation expressed by equation 2 as a population inversion, since the opposite is always true under normal conditions of thermal equilibrium. Thus, equation 2 violates the normal scheme of things. This explains why building a laser is not a trivial matter.

Laser Creation

Every laser has three basic parts: an active medium, a means of excitation, and an optical cavity. The first part requires collection of atoms that will undergo stimulated emission; the second part is the method used to achieve a population inversion; and the third part is a means of ensuring that stimulated emission exceeds spontaneous emission and results in a collimated and powerful output beam.

The active medium can be almost anything. In a CO_2 laser, it is the CO_2 molecules in a gaseous mixture of He, N_2, and CO_2. In a ruby laser, it is the Cr^{3+} ions in a matrix of Al_2O_3. In a semiconductor laser, it is the electrons and holes in a p-n junction. One means of classifying lasers is by the nature of the active medium. Some examples include a solid in a solid-state laser and a gas in a gas laser. It is probably a truism that scientists can make anything an active medium (made to lase) with the appropriate means of excitation.

The means of excitation is the source of energy that causes most of the atoms in the active medium to reach the excited state from which they can emit photons through stimulated emission. To some extent, it is possible to classify lasers according to the means of excitation. In an electrical laser, for example, an electrical discharge is the active medium that causes electrons to collide with the molecules and excites them to the upper laser level. In a gas-dynamic laser the active medium flows through a nozzle and redistributes the relative populations. In a chemical laser, the use of an exothermic chemical reaction creates the active medium in the upper laser level. In many cases, however, we may not be able to categorize so easily the means of excitation. For example, the high-temperature gases that flow in a gas-dynamic laser result from combustion of CO, O_2, and N_2, which an electrical charge may initiate.

In its simplest form, the optical cavity is a pair of mirrors designed to keep the emitted photons in line and to provide feedback into the active medium (see fig. 9-10). The optical cavity is the volume bounded by the two mirrors. When the means of excitation brings the active medium to the upper laser level, the atoms or molecules in the medium will begin to emit photons by spontaneous emission (see fig. 9-11). Most of the photons, such as A, B, and C in figure 9-11, are lost from the system. A few, such as D, may be emitted along the axis of the optical cavity (see fig. 9-12). Upon encountering mirror two, it will reflect D back into the cavity where, by stimulated emission, another photon, D′,

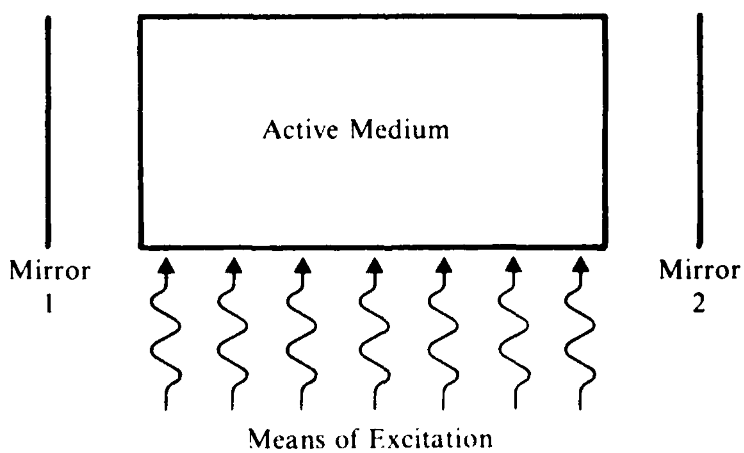


Figure 9-10. Optical cavity.

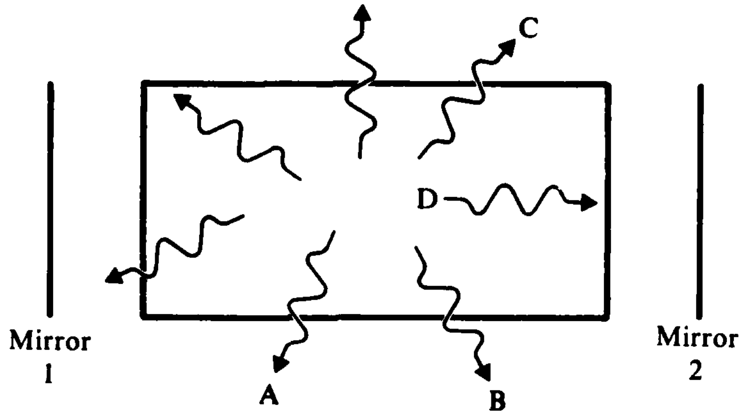


Figure 9-11. Optical cavity with spontaneous emission.

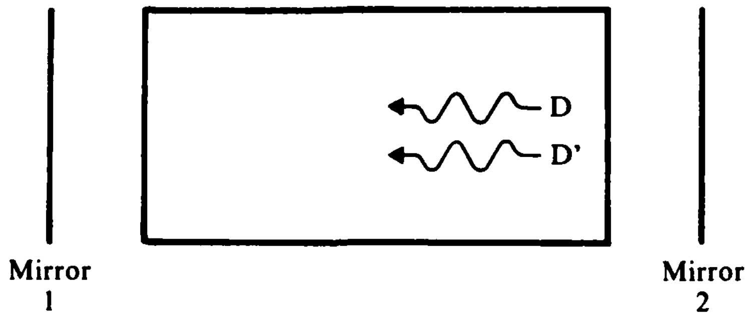


Figure 9-12. Optical cavity with feedback.

will emerge in step with D. D and D′ can continue to generate more photons, since mirror one feeds them back into the cavity. This process results in a large buildup of photons running parallel to the axis of the cavity. Of course, a practical laser needs some output. Thus, in real devices, one of the mirrors is partially transmitting so that only a few of the photons are fed back to continue stimulated emission. Photons that escape comprise the actual laser beam.

Creation of a population inversion is no small task, since all systems in nature tend to relax to the thermal equilibrium. Therefore, to make a population inversion, we must excite the atoms or molecules of the active medium into the upper laser level at a rate greater than they can decay to lower levels. What are some of the ways that excited molecules can decay to lower levels? One way is the spontaneous emission of photons. Obviously, the rate of spontaneous emission will increase if the energy difference between the two levels increases, since a population inversion more strongly violates the natural order of things as $\Delta\epsilon$ goes up. Indeed, the spontaneous emission rate A from one level to another can be related to the stimulated emission rate B between these levels by

$$A = \frac{8\,\pi(\Delta\epsilon)^3}{h^2c^3}B. \tag{3}$$

This explains why, traditionally, lasers have advanced from lower to higher frequencies. At the microwave frequencies, where laser action was first observed, excessive spontaneous emission is

not a major problem, since $\nu = \Delta\epsilon/h$ is relatively low. At visible frequencies, we need some cleverness to excite the upper laser level faster than it can decay, and, at very high frequencies (such as an X-ray laser), the size of A is a critical problem.

Molecules can lose their excitation when they collide with some other particles and give up their energy. Figure 9-13 illustrates some of the possibilities along these lines. Such losses in excitation as these depend directly on the collision of two particles. Since the density of particles increases with pressure, these losses become more important in gas lasers at high pressure. This explains why the first gas lasers operated at low pressure (< 0.1 atmosphere), and as scientists have become more adept in overcoming such difficulties, operation under higher pressure has been possible (up to several atmospheres).

Once we have achieved population inversion, the optical cavity will amplify the light as it passes through the cavity. Since the number of photons emitted by stimulated emission is proportional to the number of photons already present to stimulate the emission,

$$\frac{dn}{dt} = \propto n \tag{4}$$

where n is the density of photons in the cavity and $\propto$ the gain, which is proportional to the degree of population inversion. However, it is difficult to measure photon density (photons/m^3) and much easier to measure light intensity (watts/m^2). Therefore, equation 4 is cast normally in the form of an expression for intensity. If there are n photons/m^3 the number that will pass 1 m^2 of surface through the cavity in time Δt is $mc\omega t$ (see fig. 9-14), with c being the speed of light. Since each photon carries energy $\hbar\omega$, the intensity ($W/m^2 = d\,\frac{\text{energy}}{\text{area} \cdot \text{time}}$) traveling through the cavity at that point is thus:

$$S = \hbar\omega nc \ . \tag{5}$$

Therefore, we may write equation 4 as $\frac{d}{dt}\left(\frac{S}{\hbar\omega c}\right) = \left(\frac{S}{\hbar\omega c}\right)$. In time dt, the light travels a distance dz = cdt through the cavity and leads to the expression

$$\frac{dS}{dz} = \left(\frac{\propto}{c}\right) S = gS \tag{6}$$

where $g = \propto/c$ is the gain per unit path length in the cavity. Equation 6 explains how light builds up in passing through the cavity and has the solution

$$S(z_2) = S(z_1)\, e^{g(z_2 - z_1)} \ . \tag{7}$$

Unfortunately, a practical cavity must let some light out and thus has not only gain but loss. For the laser to work, the gain must exceed the loss. The loss is characterized in terms of a cavity "Q", or quality, defined by

$$Q = 2\pi \frac{(\text{Energy in cavity})}{(\text{Energy lost per cycle})} \ . \tag{8}$$

We may combine equations 7 and 8 to find an expression for the threshold gain, g_t, at which gain equals loss and the laser can begin to operate. If the length of the active medium is L, then one cycle in the optical cavity is the time it takes the photons to make one round trip, or 2 L/c (see fig. 9-15). If the gain is sufficient to balance losses, then the energy in cavity is simply

$$\epsilon = n_0 \hbar\omega LA \tag{9}$$

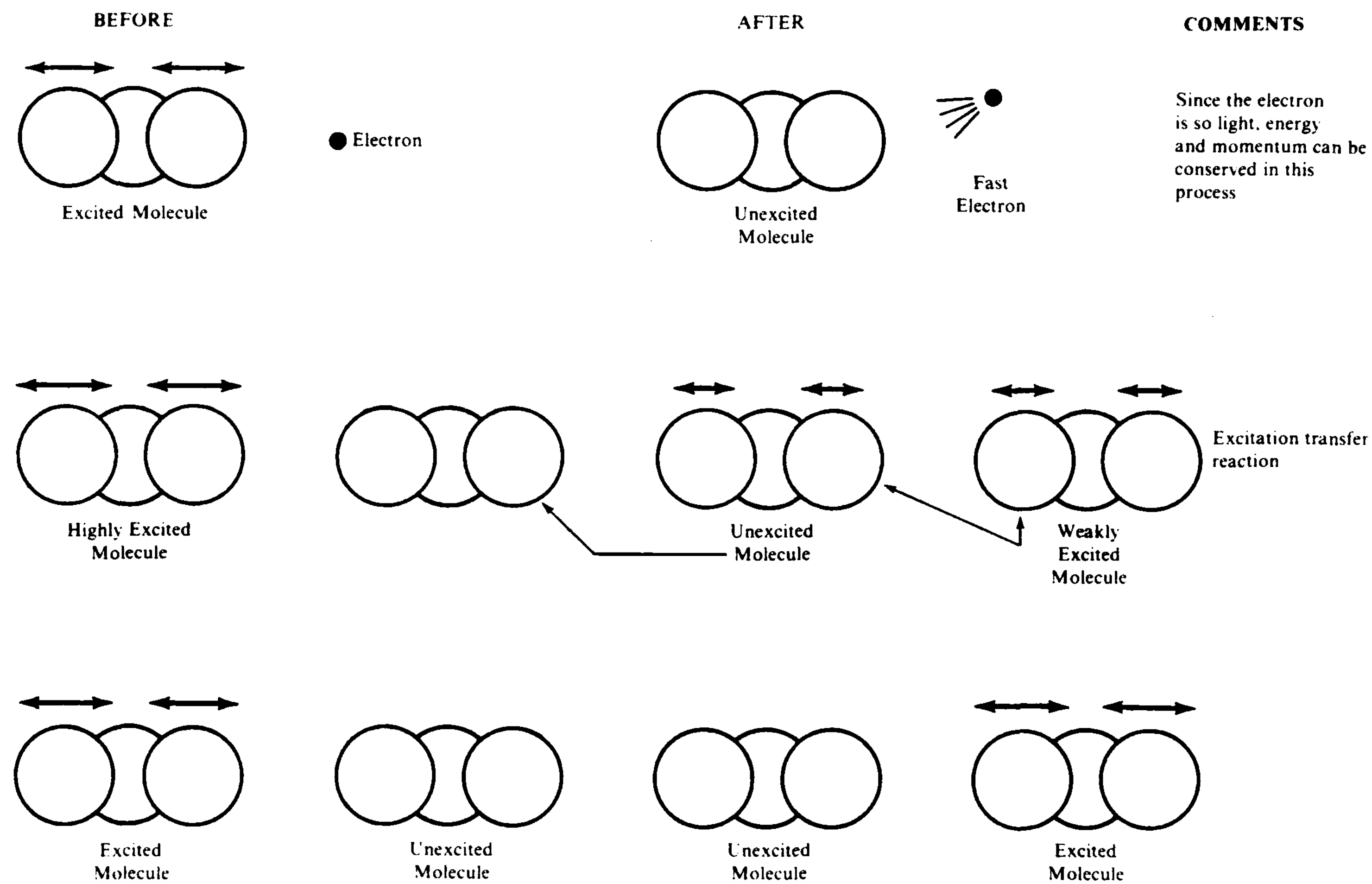


Figure 9-13. Excitation transfer processes.

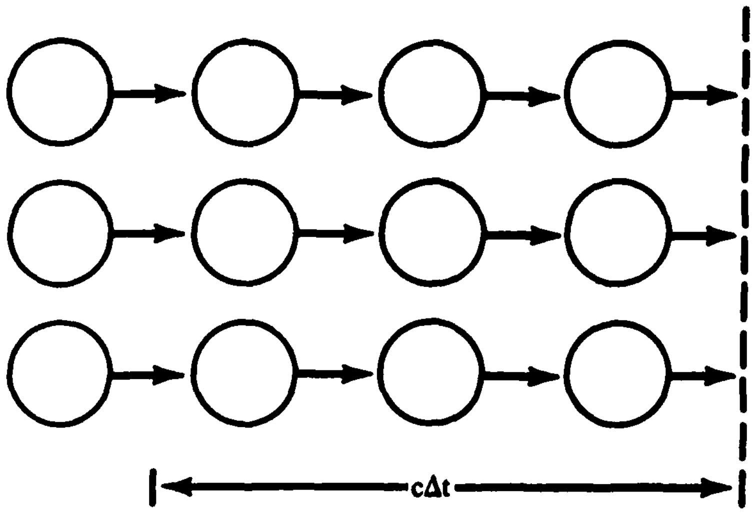


Figure 9-14. Photon Flux.

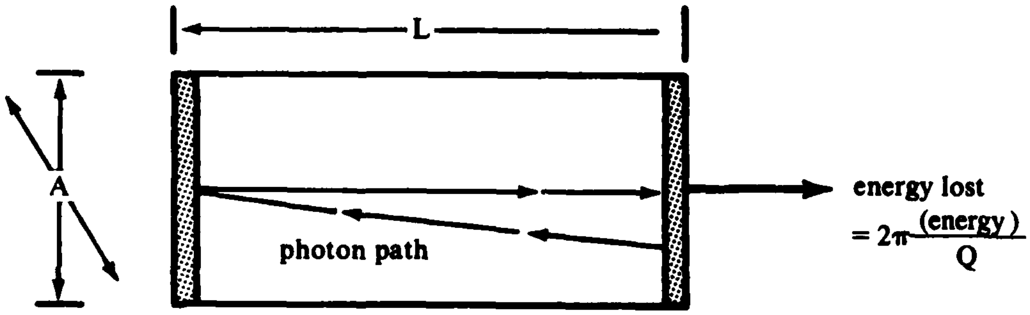


Figure 9-15. Cavity losses.

where n_0 is the photon density and A the cross section of the cavity. The energy lost in one cycle of the cavity is, from equation 8:

$$\Delta\epsilon^- = 2\pi\epsilon/Q \ . \qquad (10)$$

At the same time, the number of photons increases through stimulated emission to $n = n_0 e^{2gL}$ during one cavity cycle, so that the energy gained is

$$\Delta\epsilon^+ = \hbar\omega(n-n_0)\, LA = \hbar\omega n_0(e^{2gL}-1)\, LA \ . \qquad (11)$$

Therefore, we may find the threshold gain at which compensation occurs for cavity losses from $\hbar\omega n_0(e^{2g_t L}-1)\, LA = 2\pi n_0\hbar\omega LA/Q$ which leads to the result

$$g_t = \frac{1}{2L}\,\mathrm{Ln}\,(1 + 2\pi/Q) \ . \qquad (12)$$

Notice that as Q goes up, g_t goes down. Indeed, in the limit of large Q, $g_t \simeq \pi / QL_\theta$. Therefore, it is possible to tailor the output of a laser by experimenting with Q in a procedure known as Q-switching.

Production of a very high-intensity output beam requires a large population inversion. However, g increases as the population inversion increases, and when g exceeds g_t, lasing will begin and tend to decrease the population inversion. Thus, a low Q, or very little feedback into the optical cavity, is necessary if we desire a large g. If possible a rapid Q-switch from a low to a high value will allow buildup to a high g and rapid switch to a situation where $g_t << g$. In such a case, lasing will proceed rapidly and intensely. For example, a ruby laser without Q-switching may produce an output power of 6 kilowatts for 20 μsec, but, with Q-switching, the output may go to 600 kilowatts for 0.2 μsec. The energy produced (10^{-1} joules) is about the same in both cases. The fact that Q-switching alters the shape of the pulse may be important for high-power applications.

Overcoming the problems involved in achieving a population inversion and in achieving sufficient gain for lasing to occur will not lead to a useful high-energy device unless the optical cavity, or resonator, is designed properly. In general, we may classify resonators as either stable or unstable. Basically, stable resonators trap the light between the mirrors and permit it to get out only if one of them is partially transmitting. The stable resonator shown in figure 9-16 tends to produce a thin output beam, and operate in a few low-order radial modes.

Unstable resonators permit the light to reflect out of the cavity by going beyond the edges of one of the mirrors. The unstable resonator shown in figure 9-17 tends to produce a fat output beam, and operates in a large number of radial modes. Scientists prefer resonators for high-power applications.

Finally, an optical resonator does not need to be as simple as the resonators discussed here. In a ring laser, for example, the light travels in a circular path through the gain medium, rather than through a back-and-forth path (see fig. 9-18).

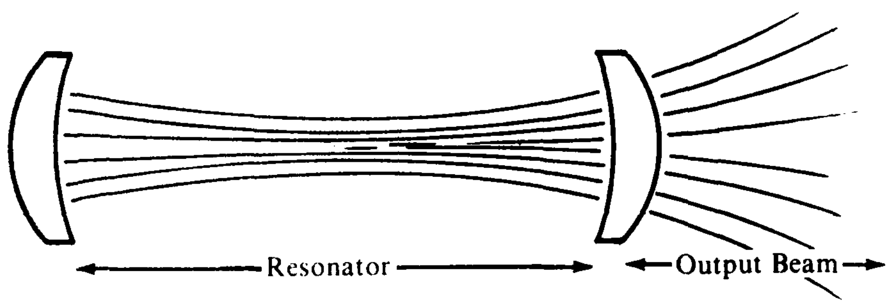


Figure 9-16. Stable resonator.

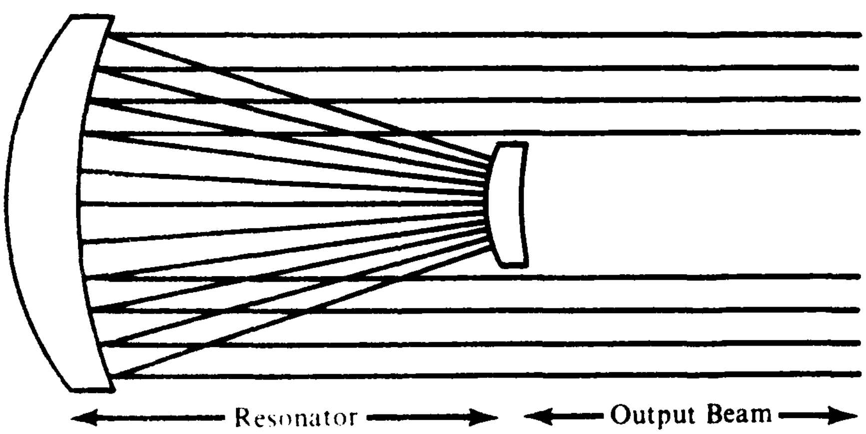


Figure 9-17. Unstable resonator.

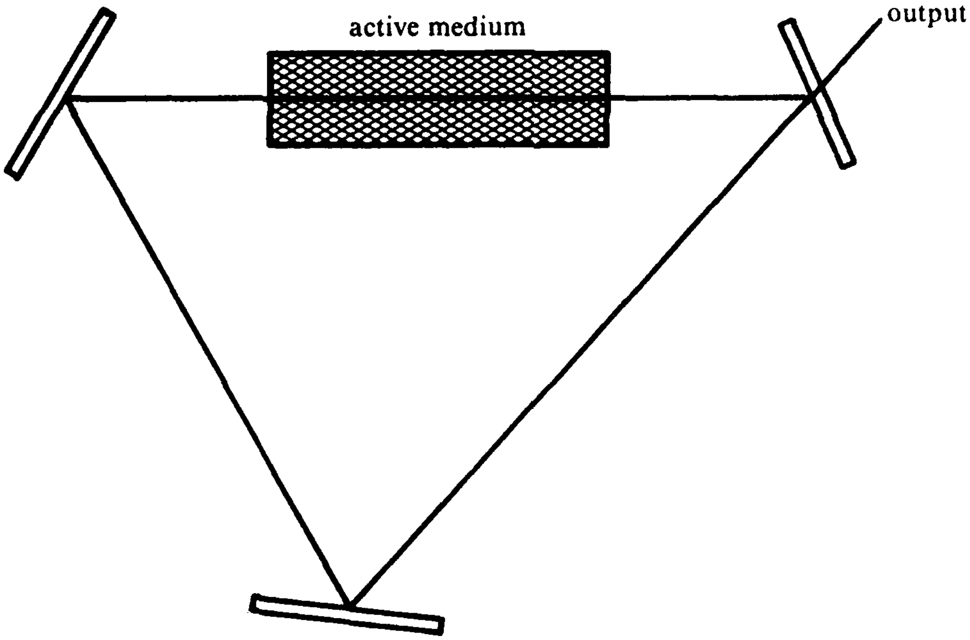


Figure 9-18. Ring laser.

Beam Propagation

Another important consideration is the way the beam propagates after it emerges. This is not so important to the laser designer, but it is important to the laser user, since propagation characteristics affect design requirements. The spot size w of a Hermit-Gaussian laser beam varies according to

$$w(z) = w_o \left[1 + \left(\frac{\lambda z}{\pi w_o^2}\right)^2\right]^{1/2} \tag{13}$$

where w_o is the minimum beam size, usually somewhere in or near the laser cavity, and z is the distance from the point of minimum spot size. Equation 13 suggests that a laser will apparently have an essentially constant w and will be a collimated beam as long as $z \leq \pi w_o^2/\lambda$, while $z > \pi w_o^2/\lambda$, it will spread in proportion to z. The quantity, $z_R = \pi w_o^2/\lambda$ which laser physicists know as the Rayleigh Range, serves as a measure of the distance the beam can travel without spreading.

Suppose we want to send a laser beam, with little divergence, to a target 1 kilometer away, and we have the choice of using a ruby laser, which operates in the visible frequency at a wave length $\lambda = 0.69\ \mu m$, or a carbon dioxide laser, which operates in the far IR at $\lambda = 10.6\ \mu m$. For a given λ and Z_R, w_o is specified. This means a minimum beam size is necessary to propagate the given distance without spreading. Comparing these two lasers we get the following chart:

λ	z_R	$w_o = (\lambda z_R/\pi)^{1/2}$	πw_o^2
0.69 μm	1 km	1.5 cm	7 cm^2
10.6 μm	1 km	5.9 cm	109 cm^2

In this example, the physics of light propagation requires that the area of a $\lambda = 10.6\ \mu m$ laser beam be $\simeq$ 16 times greater than the area of a 0.69 μm beam if both are to propagate the same distance without spread. This means if both lasers produce the same energy, the ruby laser will be approxi-

mately 16 times more effective in a welding or drilling application, since the (energy/area) would be 16 times greater. This example shows why the Air Force is interested in high-power visible lasers with output at visible frequencies, as opposed to the present generation of high-power lasers, most of which operate in the infrared frequency.

As expressed in equation 13, the tendency for a beam to spread beyond a certain point, and for that point to depend upon the initial beam size is an example of diffraction in optics. It expresses the fact that any attempt to make a smaller, more localized beam will offer no advantage because the beam will spread sooner as it becomes smaller, if all other factors are equal.

Other conditions besides the laser beam's natural tendency to spread affect its applications in the atmosphere. One of the most important of these conditions is atmospheric absorption and scattering. As mentioned earlier the photons of light in the laser beam can be absorbed if their energy, $\hbar\omega$, connects two energy levels in the molecules that make up air. Consequently, the amount of laser light absorbed depends strongly on the wavelength of the light (see fig. 9-19). Obviously, consideration of the absorption spectrum of air is just as important as consideration of the Rayleigh Range in atmospheric applications. Most of the structure seen in the IR frequency results from the excitation of vibrational modes in the molecules of air. There is much less absorption at visible frequencies. This explains the active effort to develop lasers in this region of the electromagnetic spectrum. Of course, absorption is the opposite of stimulated emission. Thus, we should not be surprised that the intensity of light decays exponentially through absorption according to the relationship

$$S(z_2) = S(z_1)\, e^{-k(z_2 - z_1)} \tag{14}$$

where k, the absorption per unit path length, is called the absorption coefficient. Comparison of equations 14 and 7 shows that absorption is basically nothing more than negative gain or attenuation. The absorption coefficient (k) depends on the density of the absorbing medium, the molecular species it contains, and the wavelength of the laser light.

To a first approximation, beam spreading through diffraction and attenuation through absorption affect beam propagation through such a medium as the atmosphere. If we believe that the energy density (J/m^2) on a target is important in causing damage, it may be interesting to note that beam spreading increases the m^2 and attenuation reduces the J. The two effects combine to reduce the energy density on target. Thus, scientists must design lasers to generate power in excess of the power actually needed to damage a target.

What happens to the energy absorbed by the air from the laser beam? It must appear somewhere, and it does appear as an increase in the temperature of the air in the channel through which the beam passes. This increase in temperature leads to an important and subtle propagation effect known as thermal blooming or thermal lensing. Since hot air is less dense than cold air, the air through which a laser beam passes will expand and become less dense than the air that surrounds it. Schematically, a beam intensity profile like the one shown in figure 9-20 will result in the air density profile shown in figure 9-21.

What difference does this make? The reader may recall from playing with a magnifying glass that a lens like the one shown in figure 9-22 causes light to focus, or bend inward. Similarly, a lens like the one shown in figure 9-23 causes light to defocus, or bend outward. The lens shown in figure 9-23 has a density profile that resembles the profile of the air through which the beam passes. Thus we should not be surprised that, as a laser beam heats the air and the air expands, the beam spreads at a rate even greater than predicted by equation 13. This is an interesting problem in dynamic feedback. The beam affects the air through which it passes, and in response, the air affects the beam.

To avoid such a problem, scientists can keep the laser pulse short, since the air requires time to expand to the condition indicated in figure 9-21. How long does it take? Pressure disturbances travel at the speed of sound, approximately 3×10^4 cm/sec. Thus, the acoustic time, τ_a, is a useful measure of the time required for thermal blooming to occur. We can determine acoustic time by the following equation,

$$\tau_a = \frac{w}{a} \tag{15}$$

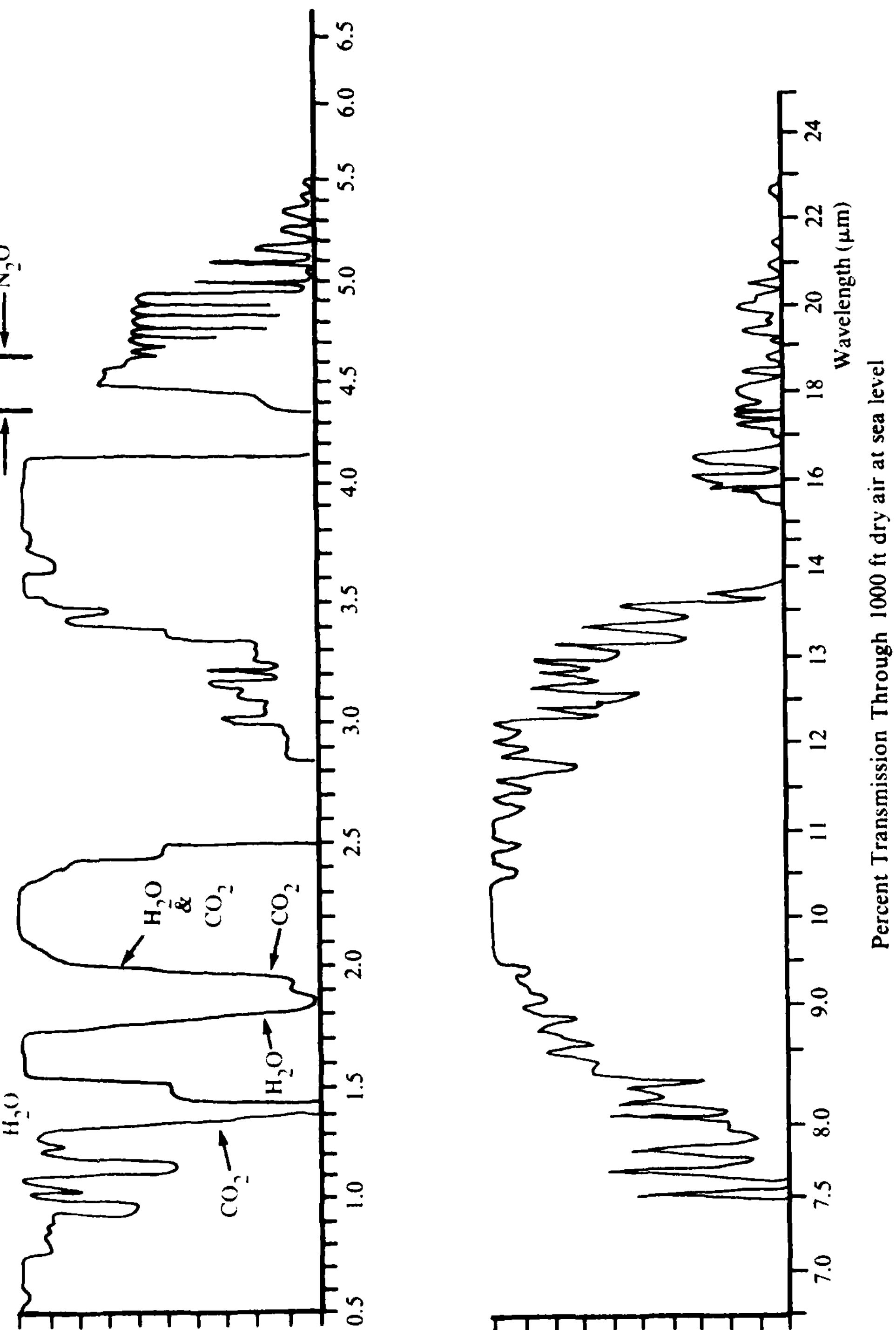


Figure 9-19. Atmospheric transmission.

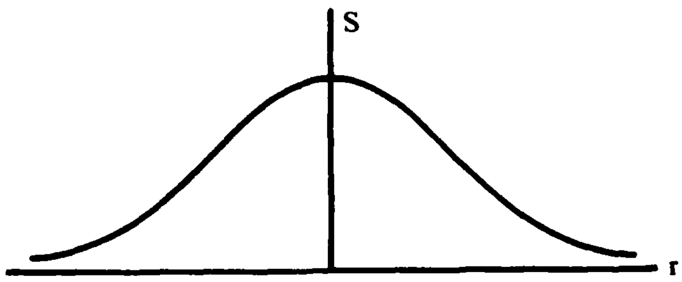


Figure 9-20. Laser intensity.

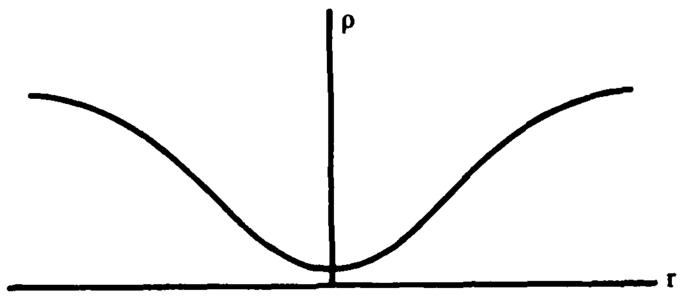


Figure 9-21. Air Density.

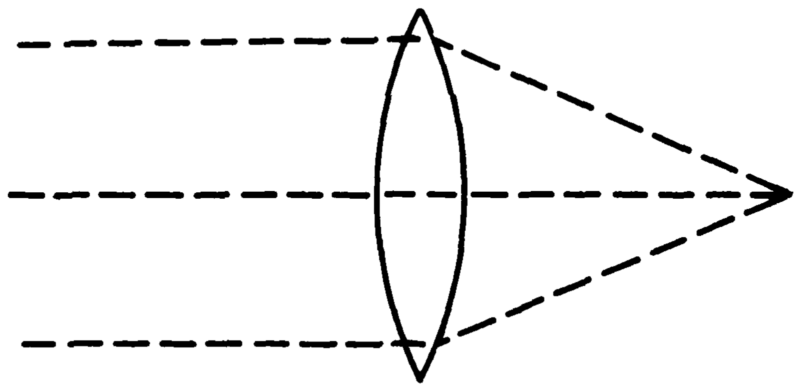


Figure 9-22. Focusing lens.

where $a = 3 \times 10^4$ cm/sec and w = beam spot radius.

This is not an exhaustive discussion of beam propagation but it should provide some insight into a few of the factors that constrain and drive device technology. The important message is that efficient propagation depends on factors that affect beam energy, size, and duration.

Summary

The following points are important to remember about lasers. The term, laser, stands for light

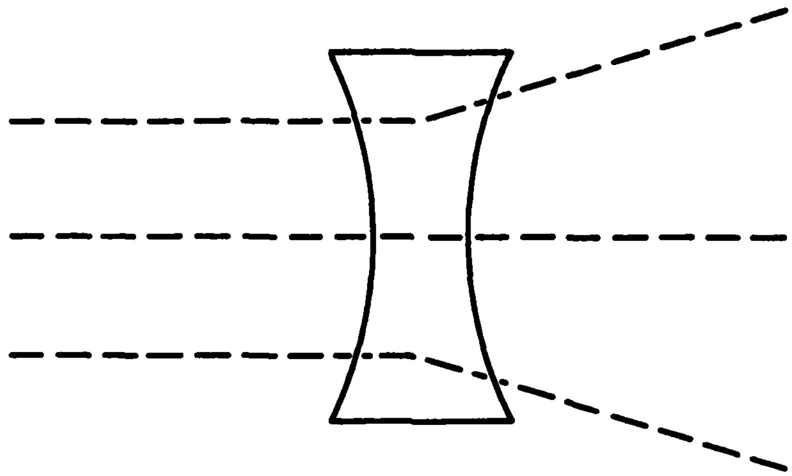


Figure 9-23. Diverging lens.

amplification by stimulated emission of radiation. Light is produced (absorbed) in transitions of atoms from higher to lower or lower to higher states of internal energy. Stimulated emission, in which light induces more light, will dominate absorption only when there is a population inversion with more excited atoms than would normally occur. Building a laser requires an active medium to emit light, a means of exciting the active medium into a state of population inversion, and an optical cavity to provide feedback for light amplification. A laser operates only at certain frequencies and may produce only specified intensity patterns. This results from cavity modes. Optical cavities may be either stable or unstable, depending on whether the light is trapped within, basically. A laser beam will propagate further with little spread or attenuation if the Rayleigh Range, $Z_r = \pi\omega_o^2/\lambda$, is long; its frequency lies outside an atmosphere absorption band; and the pulse width is chosen to minimize thermal blooming, or other nonlinear effects. With this review in mind, we are ready to study various types of lasers.

TYPES OF LASERS

We can use media in any of the three states of matter—solid, liquid, or gas—to construct lasers. We can produce population inversion in these substances in a variety of ways. These include electrical discharges, flashlamps (optical pumping), and chemical reactions.

Solid-State Lasers

The first laser, developed by Dr Theodore H. Maiman in 1960, was a solid-state device that used a synthetic pink ruby containing 0.05 percent chromium as an active medium. It consisted of a single crystal rod approximately 1 centimeter in diameter and 10 centimeters in length. The ends of the rod were mutually parallel within one minute of arc and silvered and polished to a mirror finish. One end was totally reflective, and the other end was approximately 90 percent reflective (see fig. 9-24).

A helical xenon flashtube, a device commonly used in high-speed photography, surrounded the ruby rod. The purpose of the flashtube was to generate an intense flash of light to excite the chromium atoms and cause the population inversion. The discharge of a 100 microfarad capacitor bank charged to approximately 2,000 joules (watt seconds) actuated the flashtube. The coherent radiation that emerged through the partially reflecting end of the ruby rod was in the red portion of the visible spectrum and had a wavelength of .6943 microns (note: one micron equals 10^{-6} meters). Although the ruby was the first solid-state laser material developed, it remains one of the most important

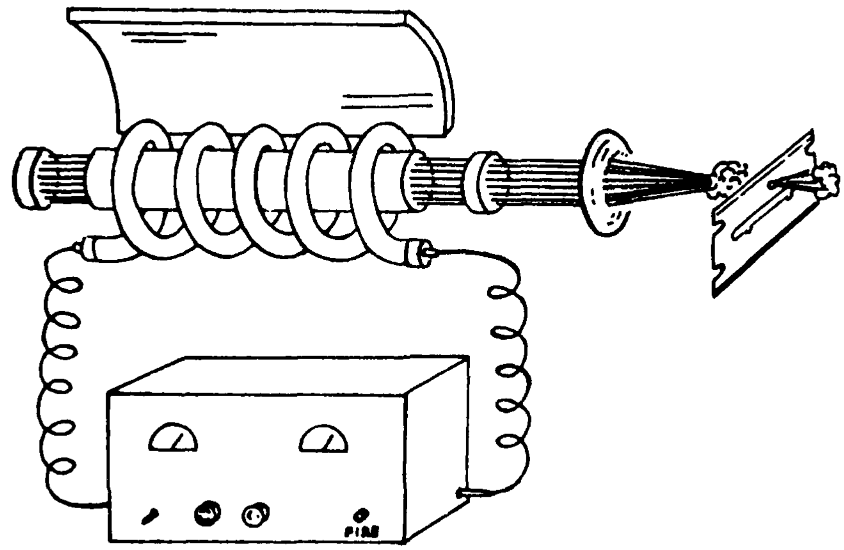


Figure 9-24. Optically-pumped solid state laser.

materials from the standpoint of power output and efficiency. Other solid-state materials are the alkaline earth tunstates and molybdates, fluorides, synthetic garnets, and optical glass laser doped with rare earth ions.

Typically the crystal, or solid-state, lasers provide high-power outputs and require powerful light sources to excite the atoms to laser threshold. Scientists have demonstrated pulsed power outputs in excess of 10,000 megawatts. However, the efficiency of present solid-state lasers is low (usually much less than 1 percent), although scientists have demonstrated some efficiencies of a high percentage. Because of the internal heating effects, we operate high-power solid-state lasers generally in a pulsed mode and cool them cryogenically. Heat causes the bandwidth of the laser to broaden, and, in some instances, it causes the laser material to stop lasing and even shatter. Another reason for pulsed operation is the practical difficulty encountered in trying to provide a sufficiently powerful continuous source of light to activate and maintain the lasing action. Typical pulse lengths of ruby lasers range from a fraction of a nanosecond (10^{-9}) to a few milliseconds. Once scientists developed the synthetic ruby laser, they were able to make hundreds of other substances, including gases, liquids, plastics, and semiconductor materials, to lase.

Gas Lasers

Shortly after scientists developed the first solid-state laser, they achieved lasing action with gas mixtures as the active medium. The configuration of an ordinary gas laser is also quite simple, consisting of a cylindrical gas discharge tube with reflecting mirrors at each end, forming an optical cavity resonator. Typically, an exciter, or radio-frequency oscillator, provides the energy to excite the gas or gas mixture and thereby activate the lasing action. Normally, we operate gas lasers at room temperatures. Argon, carbon dioxide, helium, neon, nitrogen, krypton, xenon, and various mixtures of these gases are suitable for lasing. Comparison of the coherent and monochromatic properties of the various types of lasers shows that gas lasers give the best performance. We can control their output radiation to achieve narrow bandwidths and beamwidths. Scientists have achieved spectral purities of better than one part in 10 billion in the narrowness of wavelength output. Another advantage of gas lasers is their ability to provide frequency outputs over a wide range of wavelengths. The krypton laser produces laser radiations in the red, yellow, green, and blue frequency spectrum. We can produce these radiations either simultaneously or individually. Although

the helium-neon laser is operated most often at a wavelength of .6328 microns (10^{-6} meters), we have observed more than 150 different laser transitions in the range from approximately 0.33 to 125 microns.

There are two other types of gas lasers distinguished by the excitation method used and the very high output energies obtainable. These are the gas dynamic laser (GDL) and the electric discharge laser (EDL). In a GDL (see fig. 9-25) a population inversion occurs when a carbon dioxide-nitrogen gas mixture at approximately 2,000°F expands through a nozzle. No external energy source is needed. In an EDL an electric current produces the population inversion in a gas that flows rapidly through the electric discharge region. This laser requires an electric power supply.

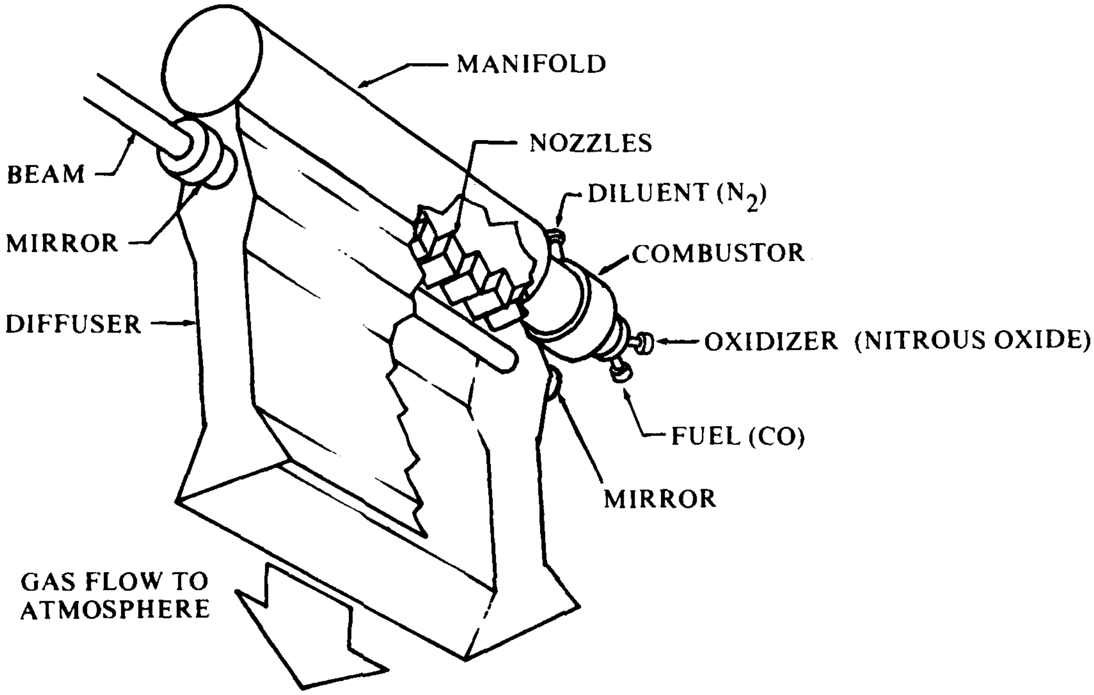


Figure 9-25. Gas dynamic laser.

Semiconductor Lasers

The simplest and most efficient type of laser is the semiconductor, or diode, laser. Although many semiconductor materials exhibit laser action, scientists have devoted the most extensive research and development to gallium arsenide (GaAs). The structure of the GaAs laser and most other semiconductor lasers is similar to the structure of any ordinary square or rectangular shaped p-n junction diode (see fig. 9-26). Small amounts of impurities introduced into the semiconductor material by diffusion processes create a positive and a negative (p-n) region. The junction between the positive and negative regions, approximately one micron thick, in most cases is the light-emitting region. Two sides of the junction region are cleaved, or cut parallel, and are highly polished to create the optical cavity. The remaining two sides are either roughly finished or offset angularly to avoid the generation of radiation in an undesired direction.

Passing an electron current through the junction region energizes this laser. Electrons passing from the n-region drop into "holes" (atoms with a deficiency of electrons) in the p-region, and emit

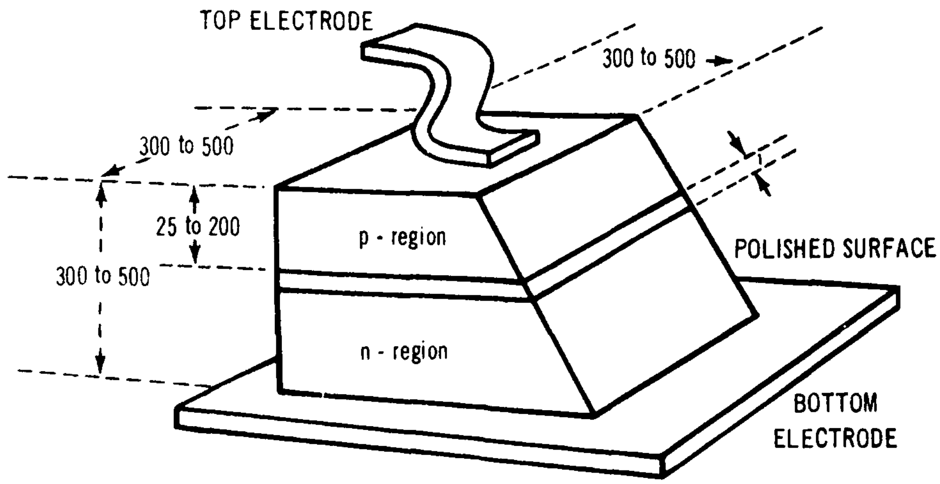


Figure 9-26. GaAs semiconductor laser (dimensions in microns).

photons in the recombination process. If the current density is sufficiently high, stimulated emission occurs.

Scientists have reported semiconductor lasers with overall efficiencies of up to 70 percent. On the other hand, these diode lasers have a small power-handling capacity because of the limited heat dissipation available in the junction. Although scientists have achieved one-watt continuous wave (CW) outputs at room temperature, cryogenic cooling becomes necessary for reliable CW operation. Placing the diode laser in contact with a metallic heat sink immersed in the coolant accomplishes the cryogenic cooling.

One major disadvantage of the semiconductor emissions is their large beamwidth, which can be as wide as 10 degrees. However, the ease of modulating the amplitude of the output beam (by varying the pumping or exciting electric current) combined with the overall high efficiency and small size of the laser makes the semiconductor laser an attractive device.

Liquid Lasers

The active species in one type of liquid laser is an organic compound dissolved in a suitable solvent. A significant advantage of using these materials is that a change of their output wavelengths can occur continuously over a wide frequency range. This provides virtually any wavelength throughout the visible frequency region as well as parts of the infrared and ultraviolet regions of the electromagnetic spectrum.

Usually the active molecules are fluorescent organic dyes (for example, Rhodamine 6G) dissolved in either water or alcohol. Flashlamps, or another laser, usually nitrogen, achieves the pumping. The selection of the proper dye molecule and the introduction of elements for selecting frequency into the optical cavity accomplish the tuning of the output radiation. Commercially available tunable dye lasers can provide continuous laser radiation throughout the visible spectrum. Automatically, the laser changes dyes at the appropriate wavelengths, and it can produce bandwidths as small as 50 megahertz.

Chemical Lasers

The active medium of a chemical laser resembles the rocket engines described in previous chapters of this handbook. Arrangement of the nozzle to produce a wide flow allows a long gain path, and the mirrors are placed transverse to the flow. A special advantage of this gas flow is removal of the

spent laser molecules from the active region as well as the excess heat generated by the chemical reaction. The hydrogen fluoride laser is typical of most high-energy chemical lasers.

The fuels in this device are molecular hydrogen (H_2) and molecular fluorine (F_2). First, the F_2 is dissociated thermally into fluorine atoms (F) in a plenum chamber prior to entrance into the rocket nozzle. Then the F and H_2 are injected into the flow through separate banks of small nozzles. As the effects of diffusion and turbulent flow cause the F and H_2 to mix, the following reaction occurs: $H_2 + F \rightarrow HF^* + H$.

HF^* represents excited hydrogen fluoride. The energy of the reaction is deposited in upper (vibrational) states of the HF molecules, and lasing occurs around 2.7 microns in the near-infrared region of the spectrum. The overall efficiency of such devices is high, amounting to several percent of the available chemical energy. The high frequency laser is capable of continuous operation at very high levels of power.

Eximer Lasers

The eximer laser is an electrically-pumped device. Electrons from an electron gun ionize atoms of a rare gas like krypton (positively) and a halogen like flourine (negatively). The ionized atoms combine to form an "excited dimer" or "eximer" that can be made to lase. Typical eximer lasers, like krypton, fluoride, and xenon chloride, lase at ultraviolet wavelengths. Eximer lasers have been around for several years. However, they have not shown the potential for either the high energy or the light weight necessary for weapon use. Scientists have studied them extensively for use in laser isotope separation.

Free Electron Lasers

The free electron laser (FEL) uses the fact that, when a beam of electrons is accelerated, that is, bent, sped up, or slowed down, it emits electromagnetic radiation in a direction tangential to the acceleration. The wavelength of the radiation emitted depends on the energy of the beam and the strength of the accelerating force. Usually, the effect is seen in synchrotron particle accelerators where strong magnets bend relativistic (very high energy) electron beams into a circle. We call the radiation produced "synchrotron radiation." This radiation can be very intense and spectrally pure, but it is only spontaneous emission. True laser radiation consists of stimulated emission produced when a photon interacts with an excited electron and stimulates it to emit a second photon identical to the first, traveling in the same direction as the first.

To build an FEL, a beam of electrons is not bent, but accelerated and decelerated in a straight line using a modulated magnetic field called a wiggler or undulator. If a laser beam with a wavelength matching the electron beam energy/magnetic field strength parameters is directed along the electron beam path, it will stimulate the electrons to emit radiation and be amplified. Alternatively, if mirrors are placed along the path of the electron beam, an optical resonator can be formed producing a self-generated laser beam. Since the wavelength is dependent only on the electron energy and the magnetic field strength, and the electrons are free and not bound to specific energy bands, theoretically, the FEL can produce or amplify any wavelength from the far infrared to the hard ultraviolet.

The first FEL amplifier was demonstrated at Stanford University in the early 1970s. Since then, scientists have directed considerable work toward the two goals of building FEL resonators and achieving short wavelengths. However, as of December 1983, only four FELs in the free world had operated with resonators. The first two ran in 1976 at long wavelengths. One was a 3 micron midinfrared laser at Stanford University and the other was a 400 micron extreme far-infrared laser built by Columbia University and the Naval Research Laboratory. The second pair ran in 1983. One was a 0.63 micron red-orange laser operated in France and the other was a 1.6 micron near-infrared laser operated at Stanford University. The pace of progress is increasing, and, in particular, French experiments have produced some encouraging surprises. However, the technology is still in its infancy. High power, beam quality, and efficiency have not been demonstrated. However, FELs hold great promise because of their inherent tunability and potential efficiency.

LASER APPLICATIONS

The most frequent potential application evoked by the word *laser* in many minds is some type of weapon or death ray. However, use of the laser as a destructive or lethal device is only one of its possible applications. Its other potential uses touch almost every field of endeavor, including communication, chemistry, biology, astrophysics, and industrial processes.

The laser has personal applications as well. For example, doctors can restore or save sight by performing a delicate laser operation called a retinal weld. This technique has become a routine operation in which photocoagulation with a suitable short but intense pulse of laser light welds the torn or injured retina to its support. The laser has shown some promise in the treatment of eye tumors and skin cancers. In fact, as a light knife, the laser may someday make bloodless surgery possible.

Laser properties have supported other fantastic accomplishments. Only two years after scientists developed the laser, a ruby laser, considerably smaller and less sophisticated than current ruby lasers, shot a series of pulses to the moon, 240,000 miles away. The laser beam illuminated a spot less than two miles in diameter and reflected back to earth with enough strength for measurement by sensitive electronic equipment. In comparison, the beam of a high-quality spotlight would spread to thousands of miles if it could reach that far. Moon-landing spacecraft used television cameras to photograph laser beams directed from earth to the moon. The spacecraft relayed the pictures back to earth. Interestingly, the laser outputs were less than three watts, illustrating the modest power required by a laser beam.

Obviously, we cannot include a discussion of all current and potential laser applications in this handbook. Therefore, we have confined our treatment of lasers to major military and space applications. Fields of particular interest to the military are communication, laser radar systems, surveillance, instrumentation, and weaponry.

Communications

The laser has major potential for use in the field of communication. The amount of information that we can encode (modulate) on an electromagnetic carrier is directly proportional to the frequency of the carrier wave. Because of the extremely high frequencies generated by lasers, they have an enormous potential capacity as transmitters of information. For example, a bandwidth of one operation on a 10^{15} hertz laser carrier wave provides an information bandwidth of 10,000 gigahertz. This information bandwidth is several orders of magnitude wider than the total frequency spectrum of the radio.

Another significant advantage is the coherence of the laser beam. This equates to high-antenna gains and permits high-data-rate communications over vast distances. Laser techniques are apparently within the state of the art for use in obtaining near real-time television pictures at interplanetary distances of 100 million miles.

Because of the spectral purity of laser radiations, effective optical filter techniques can provide a high degree of discrimination against natural background radiations from the sun, stars, and planetary bodies. The highly-collimated laser beam provides the capability of generating jam resistant, secure communication. Those trying to intercept a transmitted laser communications link can intercept only if they operate on a direct line joining the transmitter and receiver.

Laser communications offer a possible solution to the problem of radio blackout encountered by reentering spacecraft. Already, scientists have demonstrated in laboratory experiments, that, unlike radio waves, laser radiations will penetrate hot ionized gases or plasmas equivalent to reentry plasmas.

Use of the laser for communications is subject to several limitations. Laser radiations are vulnerable to absorption and dispersion by atmospheric particles, particularly during periods of rain, snow, or fog. Combined with line-of-sight constraints these atmospheric effects limit terrestrial point-to-point communications to distances of some 30 to 40 miles. Only specialized links are candidates for terrestrial point-to-point circuits because major communication systems cannot tolerate random outages (even short interruptions in service cause chaos). Two examples of such communication

circuits are secure or covert communication and data links in nuclear test environments.

Although repeaters that provide amplification and refocusing can eliminate the line-of-sight problem, we cannot eliminate the atmospheric effects. However, one means of circumventing the atmospheric aberrations is to beam the laser energy in hollow pipes that are either evacuated or filled with transparent and inert gases. These pipes with laser amplifiers every 50 to 100 miles and internal reflecting surfaces to bend the beams as necessary, could span the entire country with laser communications networks similar to existing gas and oil pipelines. Several short-distance experimental systems are already in existence in the United States, France, and the Soviet Union. Atmospheric aberrations or line-of-sight limitations would not affect pure space communications because of the near vacuum of space.

Other disadvantages of laser communications include the difficulties associated with pointing the very narrow transmitted laser beam and the problems of acquisition and tracking at the receiving end. Even disregarding the relatively low efficiency of most lasers, the efficiency of the receiving devices and other modulating and demodulating subsystems is lower at laser frequencies than at radio and microwave frequencies.

Laser Radar

Numerous existing or potential laser applications use the radar principle. These applications vary from simple hand-carried range finders for surveying or battlefield ranging to sophisticated space-tracking and detection systems. Other potential functions include spacecraft signature analysis (measurement of target length, size, and shape), altimetry, automatic space rendezvous and docking, calibration of space track radars by simultaneous tracking of targets, terrain clearance, and obstacle avoidance.

In principle, laser radar closely resembles conventional microwave radar. Scientists direct pulses of laser energy toward a target, and measure the time elapsed between the pulses and their reflected returns. They determine the range directly from the product of the known speed of the laser energy (the speed of light) and the elapsed time. The difference between the two systems is due primarily to the higher laser frequency. The elements of the laser radar are merely the optical analogy of their microwave counterparts.

The range resolution of a pulsed radar system depends on the sharpness of the leading edge of the pulse and the accuracy in measuring the transit time of the two-way pulse. This highly collimated laser beam permits range measurements in environments where conventional radars fail because of their wider beams. For example, reflections from target background clutter render conventional fire control radars in low-flying aircraft useless at attack angles of less than approximately 15 degrees. Laser systems can provide tracking data during the launch and initial flight phase of space vehicles when ground clutter reduces the effectiveness of conventional radar.

In a space-tracking role near the earth, conventional radar can determine range to an accuracy of approximately 100 feet; the laser narrows the error to approximately 25 feet. With cooperative satellites equipped with special mirrored corner reflectors, accuracies are better than 10 feet at this distance. Scientists have demonstrated successful ranging on a cooperative satellite at ranges of 1,000 miles.

Laser radars were particularly effective in supplementing the Baker-Nunn cameras that were used on several satellite-tracking networks. When the spacecraft was traveling in the earth's shadow, they provided range information in addition to illumination for photography. Since the Baker-Nunn camera provided angular coordinates, the additional range information specifies the position vector of the spacecraft at the time of observation.

Another advantage of the laser is the modest optical lens aperture required to concentrate its emitted beam. By using a visible laser, a lens one foot in diameter can provide a one-minute (of arc) beam. In comparison, the FPQ-6 radar used in a space-tracking role requires an antenna 30 feet in diameter to produce a 24-minute (of arc) beam. Angular resolution is a function of the narrowness of the radiated beam. Because of the narrow beam, present target azimuth and elevation angle resolution

shows an improvement of at least an order of magnitude over conventional radars.

As in communications, however, laser radar has several disadvantages in comparison to conventional radar. These include the difficulty in acquiring targets, and for terrestrial applications, exposure to the attentuations and aberrations caused by the atmosphere, particulary under adverse weather conditions.

Surveillance

The military can use lasers for such surveillance functions as imaging, multicolor sensing, moving target indication, and illumination of covert photography. Competing technologies are passive infrared, side-looking radar, and conventional photography. Laser systems offer advantages of better resolution than passive infrared and side-looking radar and more covert night operation than conventional photography.

The covert nature of an active laser surveillance system is based on two factors: it provides its own illumination and it uses a narrow pinpoint beam in a rapid sequential scanning process. In providing coherent self-illumination, it permits operation under day and night conditions. The highly collimated radiation permits a high degree of covert position even when frequencies are used in the visible range, and use of frequencies in the visible infrared or ultraviolet spectrum would further enhance covert operations. Military forces can use the laser in such a special application as terrain profile, or contour, mapping. It could improve spacecraft signature analysis as well.

Another important area where the laser may improve present capabilities is underwater surveillance. Although scientists know the attentuation of light transmission to be several orders of mangitude greater than sound waves (sonar application), the use of laser energy in the blue-green region of the visible frequency spectrum provides hope that we can develop viable underwater surveillance systems. Seawater offers the least attenuation to these frequencies, and the sensitivity of laser radiation detection devices is highest in the same frequency range. Scientists are aiming present research at minimizing problems of backscattering due to suspended organisms and minerals, as well as water.

Instrumentation

Many interesting laser applications have been made in instrumentation. These applications include radar and gun bore sight alignment techniques, optical computers and data processing, wideband high-resolution recording, techniques for measuring vibration of spacecraft in simulated flight tests, and visual display technology, which promises to produce screen sizes at least an order of magnitude greater than present TV-type displays.

Another fascinating area is holography, a means of storing (making a hologram) and reproducing, at will, a three-dimensional image of an object or scene. We can produce holograms in black and white or color. The reconstructed, stored image is a virtual image floating in space some distance behind the hologram. If the observers change viewing positions, they literally, can look around the object or elements in the scene. The broad range of potential applications of holography almost rivals its parent field of laser applications.

Another potentially valuable use of the laser involves the laser gyro, which does not depend upon the properties of inertia. Since the laser gyro has a rugged and simple design with no moving parts including gimbals, scientists can attach it rigidly to a vehicle. Its other advantages include the capability to operate in environments of extremely high acceleration and to perform almost instantaneously after being switched on.

The laser gyro is based on the principle that rotation of the optical cavity about its axis will cause frequency differences to develop between two simultaneous contrarotating laser beams in a closed-loop optical cavity since the frequency of laser operation is a function of cavity length. If the optical path is stationary, the two contrarotating beams take the same time to traverse the loop, and their frequencies are the same. If we rotate the gyro about its axis, we shorten the effective length of the path for one beam and lengthen the path of the other. This causes the frequency of one beam to

increase and the other to decrease, and we may observe the resulting difference by optical heterodyne techniques. The magnitude of the difference is proporti ·al to the rate of rotation. Scientists can position three planar laser gyros orthogonally to provide information on pitch, roll, and yaw.

Weaponry

Shortly after the invention of the laser, it became apparent that if high-energy laser weapon systems could be developed, they would have some particularly attractive features. For example, since light travels at a speed of 186,000 miles per second, the lethal fluxes would arrive at targets almost instantaneously, and there would be no requirement to lead the targets unless they were located very great distances from the weapons. The military could use laser weapons for selective engagement of single targets in the midst of numerous friendly vehicles. Unlike nuclear weapons, laser weapons would not disturb large segments of air space and would not destroy indiscriminately all vehicles within their lethal range.

A high-energy laser weapon system is a system that attempts to inflict damage on an aerospace vehicle by placing large amounts of energy on a small area. The result is a thermal kill, such as weakening and eventual rupture of structural components, ignition or combustion of flammable materials, or destruction of thermally sensitive items in vital components.

Initially, scientists considered solid-state lasers (large ruby rods or large yttrium aluminum garnet rods known as YAG) suitable for high-energy laser applications. Indeed, they could achieve short pulses of very high power with these lasers. However, internal heating of the crystal resulted in fracture of the crystal long before it could deliver damaging amounts of energy to the target. Further development was necessary.

Scientists and engineers solved the problem in 1967 with the invention of the carbon dioxide GDL, or CO_2 GDL. The CO_2 GDL, the first GDL that appeared scalable to very high energies, has paved the way for serious consideration of a high-energy laser weapon system. In recent years, scientists have developed other gas-phase systems, including the carbon dioxide EDL, the carbon monoxide EDL, the deuterium fluoride chemical laser, the hydrogen fluoride chemical laser, and several others.

Military leaders have studied many different targets to determine their vulnerability to a hypothetical laser weapon system. At one end of the spectrum, numerous items, such as infrared sensors, are exceptionally vulnerable to laser radiation, particularly when the radiation affects the band normally accepted and processed by the sensor. At the other end of the spectrum, scientists have designed such items as ballistic missile reentry vehicles to survive very high temperature environments, thus, they are more difficult for the laser's thermal kill mechanism to damage. Numerous attractive targets appear to lie between these two extremes.

PARTICLE BEAM DESIGN AND DEVELOPMENT

Charged particle beams (CPB) are not useful as weapons in space, but an understanding of their generation and propagation in the atmosphere is still important. After all, neutral particle beams cannot be accelerated unless they first exist as CPBs in the accelerator. An understanding of the physics of charged particle beams will explain, to some degree, why they cannot propagate from the atmosphere into space, cannot propagate for any useful range as weapons in space if generated there, and cannot be accurately pointed in space because of the earth's magnetic field. This explanation shows as well why neutral particle beams cannot propagate in the atmosphere, but might achieve useful weapon ranges in space.

Basic Ideas

Usually charged particles consist of electrons with negative charges and protons with positive charges. Both have a charge of magnitude 1.6×10^{-19} coulombs, but their masses are quite different. An electron has a mass of 9.11×10^{-31} kilograms, and a proton has a mass of 1.67×10^{-27} kilograms approximately 1,000 times greater. Since the particles are charged, we can accelerate them to very

high energy with a suitable electric field. The amount of energy received by an electron during acceleration across a potential of one volt is called an electron volt (eV). It is equal to 1.6×10^{-19} joules. This convenient unit of energy for CPB problems is used almost exclusively in the literature on the subject.

Accelerating particles of a given charge in an electric field of given strength is reasonably straightforward, but some knowledge of relativistic dynamics is necessary to understand what happens to the particles at high energy. The speed of a particle cannot exceed the speed of light, and strange things happen when the particle approaches this limit. The effective mass of the particle changes. If m is the mass of the particle at rest (called rest mass), then its mass at velocity v is

$$m' = \gamma m \tag{16}$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \,. \tag{17}$$

In equation 17, we can see that $\gamma \simeq 1$ and; $m' \simeq m$ as long as $v < c$. As v approaches c, m' increases without bound.

What is the impact of this effect on the laws of motion? We can rewrite Newton's law as follows

$$\underline{F} = m\underline{a} = m\frac{d\underline{v}}{dt} \,. \tag{18}$$

This relates applied force to particle acceleration, but at relativistic velocities, m is no longer constant and we must write equation 18 in the more general form

$$F = \frac{d\underline{P}}{dt} = \frac{d}{dt}m'\underline{v} = \frac{d}{dt}\gamma m\underline{v} \tag{19}$$

where P is the particle momentum. As v gets closer and closer to c, an applied force is effective primarily in changing γ rather than v, since there is an upper boundary to how high it can go. This leads to the question of definition regarding the energy of a relativistic particle. The kinetic energy of a nonrelativistic particle is

$$T = {}^1/_2\, mv^2 \,. \tag{20}$$

The equivalent formula for a relativistic particle is

$$T = (\gamma - 1)\, mc^2 \,. \tag{21}$$

At first glance, it is difficult to see the relationship between equations 20 and 21; but at slow speed,

$$\gamma = \frac{1}{(1 - v^2/c^2)^{1/2}} \simeq 1 - {}^1/_2\frac{v^2}{c^2} \tag{22}$$

so that for $V << C$,

$$T = (\gamma - 1)\, mc^2 \simeq (1 - {}^1/_2\frac{v^2}{c^2} - 1)\, mc^2 = {}^1/_2\, mv^2 \,.$$

In other words, at nonrelativistic velocities, equation 21 does reduce to equation 20. We can define the total energy of a particle as

$$E = \gamma mc^2 \quad . \tag{23}$$

This includes both the kinetic energy of motion, $(\gamma - 1)\,mc^2$, and the "rest energy" mc^2. Rest energy is the energy of a particle when it is motionless and it is energy that we can release only if we annihilate the particle and convert it entirely to energy. The rest energy of an electron is $m_ec^2 = 0.511$ MeV $\simeq 0.5$ MeV and of a proton it is $m_pc^2 = 938$ MeV $\simeq 1$ GeV.

The electromagnetic field of a charged particle beam is an important consideration in its creation. Figure 9-27 shows an infinitely long beam of particles of charge q and mass m, all moving with velocity v. Let ω be the radius of the beam and n the assumed uniform density of particles within it. Then the charge density within the beam is

$$\rho = nq \tag{24}$$

and the current density is

$$j = nqv \quad . \tag{25}$$

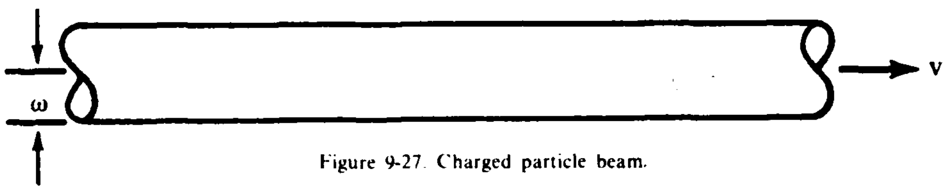


Figure 9-27. Charged particle beam.

The electric field associated with the charges in the beam has only a radial component in cylindrical coordinates, and, from div $\underline{E} = \rho/\epsilon_o$, we get

$$\frac{1}{r}\frac{\partial}{\partial r} rE_r = \frac{\rho}{\epsilon_o} = \frac{nq}{\epsilon_o} \tag{26}$$

Since ρ is constant for $r < \omega$ equation 26 may be immediately integrated to give $E_r = \frac{1}{2}\frac{nqr}{\epsilon_o}$ if $r \leqslant \omega$; $E_r = \frac{1}{2}\frac{nq\omega^2}{r\epsilon_o}$ if $r > \omega$. We have illustrated this field in figure 9-28.

Similarly, the magnetic field associated with the moving charges has only 0 − component, and from curl $\underline{B} = j/\epsilon_oc^2$, we get

$$\frac{1}{r}\frac{\partial}{\partial r}(rB_o) = -\frac{nqv}{\epsilon_oc^2} \tag{27}$$

In structure, equation 27 is the same as equation 26, thus, we get

$$\begin{aligned} B_o &= -\frac{1}{2}\frac{nqvr}{\epsilon_oc^2} \text{ if } r \leq \omega \\ &= -\frac{1}{2}\frac{nqv\omega^2}{r\epsilon_oc^2} \text{ if } r > \omega \end{aligned}$$

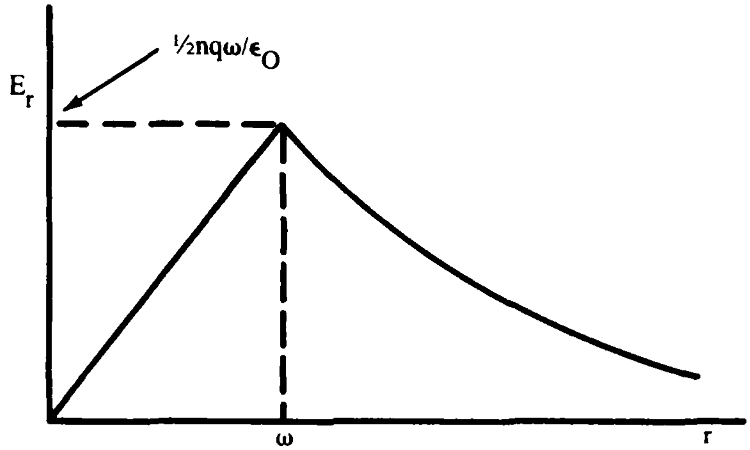


Figure 9-28. Electric field of a CPB.

E_r and B_θ are proportional; in fact, $B_\theta = \frac{v}{c^2}E_r$. Thus, we can find the force felt by a particle in the beam as a result of the fields generated by all the other particles from

$$\underline{F} = q\,(\underline{E} + \underline{v} \times \underline{B}) \quad . \tag{28}$$

Since v is in the z-direction, $\underline{F}$ has only a radial component given by

$$F_r = q(E_r - vB_\theta) = q(E_r - \frac{v^2}{c^2}E_r) = q\,E_r(1 - \frac{v^2}{c^2}) = q\,\frac{E_r}{\gamma^2} \quad . \tag{29}$$

In other words, a net force outward decreases in magnitude as γ goes up. This explains why scientists prefer relativistic particles for particle beam applications.

Scientists and engineers who work with charged particle beams prefer to characterize them in terms of particle energy and beam current rather than particle density and velocity. The beam current in amperes is simply

$$I = \pi\omega^2\,qnv \tag{30}$$

and the velocity v is

$$v = c(1 - \frac{m^2c^4}{\epsilon^2})^{1/2} \quad . \tag{31}$$

In terms of the beam current, then, the magnetic field associated with the beam is

$$B_\theta = -\frac{1}{2\pi\omega^2}\frac{Ir}{\epsilon_0 c^2} \quad r \leq \omega$$

$$= -\tfrac{1}{2}\,\frac{I}{r_0 c} \quad r > \omega \tag{32}$$

and the electrical field is

$$E_r = \frac{c}{(1 - \frac{m^2c^4}{\epsilon^2})^{1/2}}\,\frac{Ir}{2\pi\omega^2\epsilon_0 c^2} \quad r \leq \omega$$

$$= \frac{c}{(1 - \frac{m^2c^4}{\epsilon^2})^{1/2}}\,\frac{I}{2\pi r\epsilon_0 c^2} \quad r < \omega \tag{33}$$

What are some typical values? If we have a 1 kAmp, 1 GeV electron beam of radius 1 cm = 10^{-2} m, then the magnitude of the magnetic field at the edge of the beam (r = ω) is

$$B_\theta(\omega) = \frac{1}{2\pi} \frac{(10^3\,\text{Amp})}{(10^{-2}\,\text{m})(8.85\times 10^{-12}\,\text{fd/m})(3\times 10^8\,\text{m/sec})^2} = 2\times 10^{-2}\,\text{Weber/m}^2 = 200\ \text{Gauss}$$

and the electric field is

$$\begin{aligned} Er(\omega) &= \frac{c}{\left(1 - \frac{m^2c^4}{\epsilon^2}\right)^{1/2}} B_\theta(\omega) \\ &= \frac{(3\times 10^8\,\text{m/sec})}{\left[1 - \left(\frac{0.5\times 10^6\,\text{ev}}{10^9\,\text{ev}}\right)^2\right]^{1/2}} (2\times 10^{-2}\ \text{Weber/m}^2) \\ &= 6\times 10^6\ \text{v/m}\ . \end{aligned}$$

These are fairly substantial fields, a point that we will consider again in a later discussion.

Beam Creation

An accelerator is the device that creates a CPB. Like lasers, all accelerators have certain elements in common. These elements include a source of particles and a method of injection into the accelerator, a strong electric field to accelerate the particles, some method of handling the beam during the process of acceleration, and a method of extracting the beam into the outside world. Accelerators come in three varieties: linear, circular, and "collective effect."

Linear accelerators are the devices that most people would produce if asked to design accelerators. In its simplest form, a linear accelerator looks similar to the sketch in figure 9-29. The particle source could be some radioactive material that emits charged particles. The particles enter into the acceleration region, and, at some time, a potential v is applied across the region. The accelerated particles emerge through a hole in the other end of the device.

Such a simple scheme poses a number of problems. In the first place, it depends on some fairly strong electric fields to achieve the required energies. In practical terms, E-fields of greater than 1 MV/m are difficult to achieve because of a tendency toward field emission from the plates, arcing, and breakdown. If we assume that we can achieve the limit, then the length of the accelera-

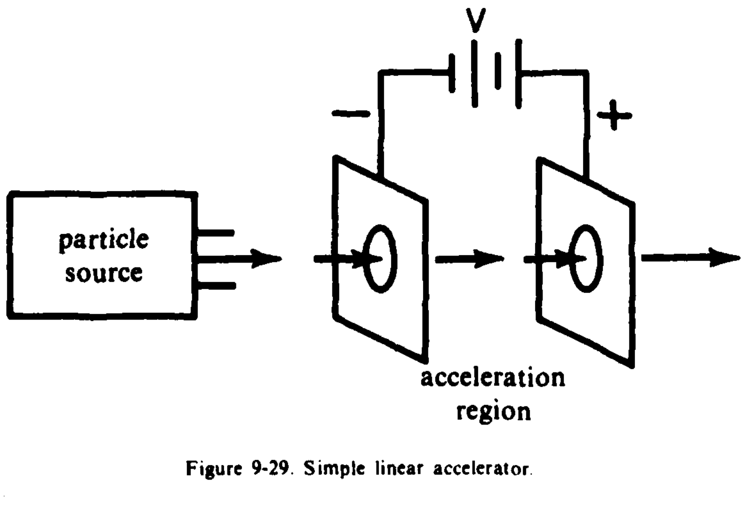


Figure 9-29. Simple linear accelerator.

tion region necessary to achieve 1 GeV energy is 1 kilometer. Capacitor plates large enough to maintain a uniform field for such distance are difficult to make. Practical linear accelerators (sometimes called linacs) appear more like the sketch shown in figure 9-30.

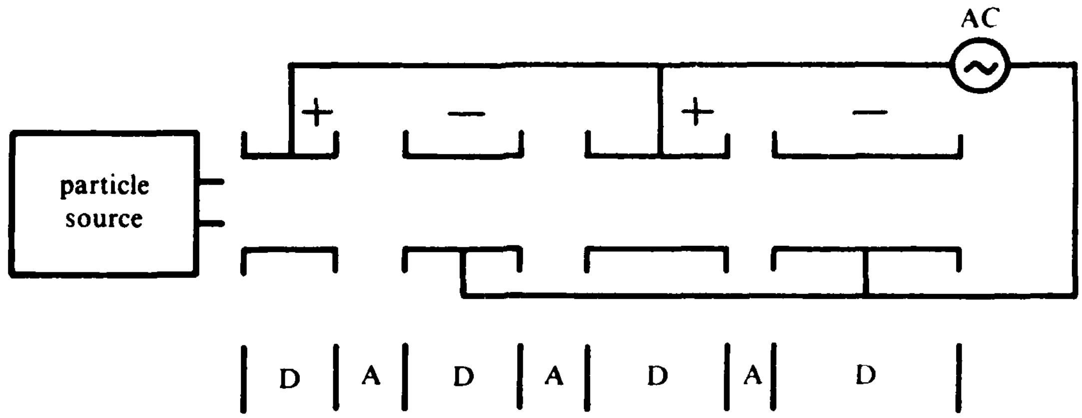


Figure 9-30. A more practical linear accelerator.

This accelerator consists of several acceleration regions (A) broken up by drift regions (D). We apply an AC voltage in a frequency that reverses the sign of the E-field while the particles are in the drift regions. Consequently, they are always accelerated in the forward direction and the length of the drift regions becomes greater as the particles go faster. This arrangement overcomes the problem of maintaining a uniform E-field, but it does nothing for total length. Indeed, the linear accelerator at Stanford is more than two miles long!

Circular accelerators offer a means of making accelerators of more modest size. Figure 9-31 shows one such accelerator, called a cyclotron. The cyclotron operates as follows. The drift regions are now semicircles, and we must apply a strong magnetic field to make the particles move in a circular path. Again, we apply an E-field to the gap, and choose its frequency to provide acceleration while the particles are in the gap. The particles are injected in the center, and, as they gain energy, the radius of their orbit increases until they are skimmed off at the edge. Because of the circular path, the particles travel a long way ($\sim$50,000 miles) but the accelerator itself can fit in a room of modest size.

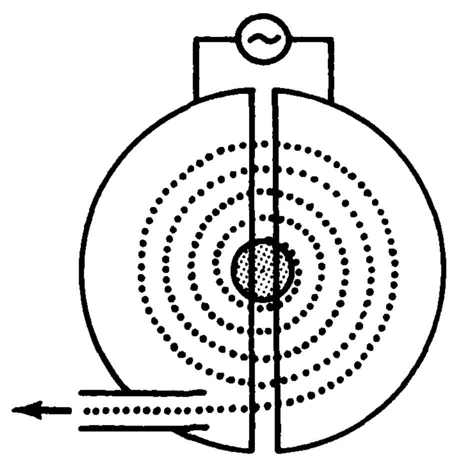


Figure 9-31. Cyclotron.

Comparison of linear and circular accelerators shows that linear accelerators offer advantages in particle injection and extraction, and circular accelerators offer compactness. However, both present problems in dealing with high-energy, high-current beams. As mentioned earlier, the E-field at the edge of 1 GeV, 1 kAmp CPB is approximately 6×10^6 v/m. On the other hand, the maximum E-field in the acceleration region might be something like 1×10^6 v/m. Therefore, the forces tending to cause the beam to spread radially are greater than the forces pushing it forward in the accelerator. Consequently, achievement of both high current and high energy has been beyond the state of the art, since it would require a large charge density within the beam and strong radial forces as indicated in equations 32 and 33. Tabulated below are some examples of present day accelerator parameters.

TYPE	*ENERGY (MeV)*	*CURRENT (Amp)*
High-ϵ, Low I		
Oak Ridge	178	20
Stanford	1,200	3×10^{-2}
Low-ϵ, High I		
University of Maryland	5	2×10^4
MIT	1.6	2×10^4

Collective-effect accelerators are a possible means of overcoming these difficulties. The theory is to use the strong E-field associated with a high-current beam rather than allow it to hinder. One concept works in this manner. We start with a high-current, low-energy electron beam, since this is easily within the state of the art (fig. 9-32). We make waves on the beam. The waves might be fluctuations in charge density produced by an AC voltage applied to a grid through which the beam passes (fig. 9-33). Now the density fluctuations in the beam induce a periodic variation of electric potential in the z-direction (fig. 9-34). If protons were introduced into the beam, they could become trapped in the potential wells and carried along with them. The result would be protons moving at the same velocity as the electrons. However, the velocity of relatively low-energy electrons ($\sim$1 MeV) is the same as the velocity of high-energy protons ($\sim$1 GeV). Thus, the idea holds out the hope of a compact, high-energy, high-current proton accelerator.

The idea may appear farfetched at first glance, since no one has built a working accelerator along these lines as yet. But the principle is not as farfetched as it might appear. Surfers regularly use water waves to accelerate themselves in an analogous process. Two obvious questions come to mind: What is the source of the protons and how can they be trapped in the waves? One way to introduce protons into the beam might be to create them by sending the beam through a gas of hydrogen, which it would ionize. Trapping the protons in the waves would be a more difficult problem. It would require slowing the wave velocity (not necessarily the beam velocity) down to a velocity approaching zero and then speeding it up again to accelerate the protons. The government is spending a fair amount of research and development money to investigate these problems.

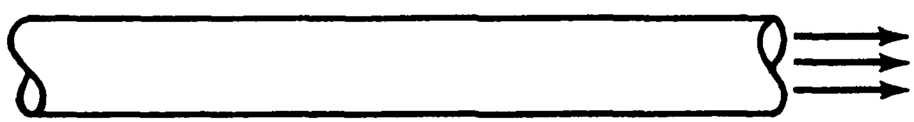


Figure 9-32. Electron beam.

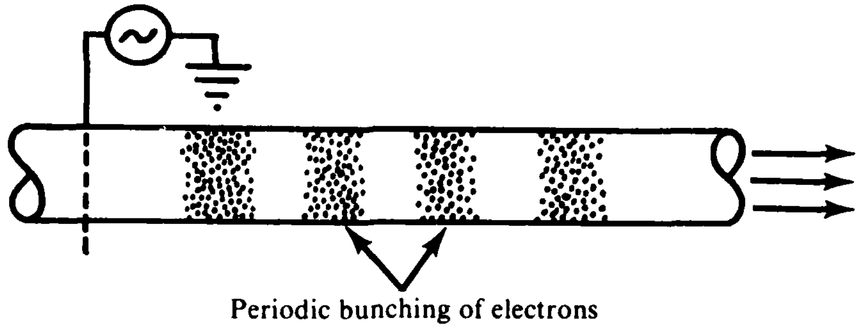


Figure 9-33. Electron beam with waves.

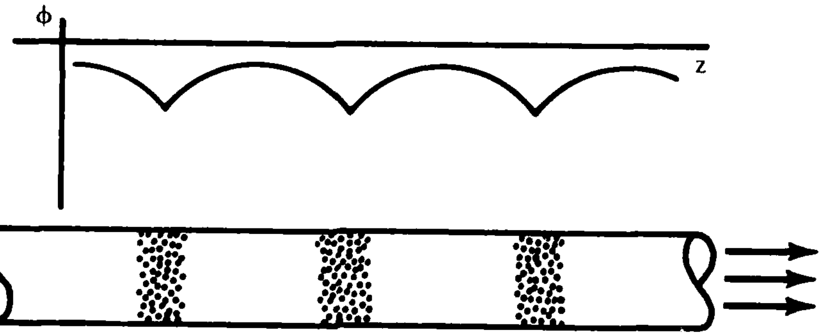


Figure 9-34. Electric potential of an electron beam with waves.

Beam Propagation

The study of laser propagation was fairly straightforward. Items considered were the natural tendency of the beam to spread at distances beyond the Rayleigh Range, absorption into the air, and such nonlinear effects as thermal blooming. CPB propagation is more complex since almost everything discussed in this area is pure theory. As yet, scientists have not made high-current accelerators to test whether the beam will propagate according to projections.

Beam spreading, the first problem, is much more severe than beam spreading in the case of lasers. As equation 29 shows, a net force pushes the beam outward, given by $F_r = qE_r/\gamma^2$, and a typical value for E_r is roughly 10^7 v/m for beams of both high energy and high current. For a 1 GeV electron beam, $\gamma = 1{,}000$; therefore,

$$F_r = \frac{(1.6 \times 10^{-19}\ \text{coul})(10^7\ \text{v/m})}{(10^3)^2} = 1.6 \times 10^{-18}\ \text{nt}$$

Using Newton's law (equation 19), the radical acceleration felt by the particle in the beam is

$$\frac{dv_r}{dt} = \frac{d^2r}{dt^2} = \frac{1}{\gamma m}F_r = \frac{1.6 \times 10^{-18}\,\text{nt}}{10^3(9.1 \times 10^{-31}\ \text{kg})} = 1.76 \times 10^9\,\text{m/sec}^2$$

Since v_r in this case is initially zero and the acceleration is perpendicular to the direction of motion,

$$\frac{d}{dt}(\gamma m v_r) = \gamma m \frac{dv_r}{dt} .$$

Therefore, we can estimate the time it takes a beam with a 1 centimeter radius to double in size as

$$t_d = \left(\frac{\omega}{d^2r/dt^2}\right)^{1/2} = \left(\frac{10^{-2}\ \mathrm{m}}{1.76 \times 10^{9}\ \mathrm{m/sec^2}}\right)^{1/2} = 2.5 \times 10^{-6}\ \mathrm{sec}\quad .$$

During this time, the beam will travel a distance

$$z_d = v\tau_d \simeq c\tau_d = (3 \times 10^{8}\ \mathrm{m/sec})(2.5 \times 10^{-6}\ \mathrm{sec}) = 750\ \mathrm{m}.$$

Therefore, we cannot expect an electron beam to go more than a kilometer without doubling in size and reducing its intensity (W/m^2) by a factor of four. For protons, the situation is worse. In that case, m_p is greater than m_e, but $\gamma \simeq 2$ and F_r and dv_r/dt are also greater, giving a double distance of $Z_d \simeq 3$m.

Based on these simple examples, we might conclude that charged particle beams are not viable candidates as weapons, since any realistic application would require much greater ranges. In fact, such a conclusion would be correct in outer space where there would be nothing to limit the beam from spreading under its own internal forces.

For applications in the air, ionization may be a means of avoiding the spreading tendency of the beam. As the ionizing process proceeds, the beam soon finds itself propagating, not through neutral air, but through a channel of ionized gas having high conductivity. Under these circumstances, charges will flow in such a way that the E-field is neutralized on the interior, leaving the imbalance in the net charge only on the surface. Figure 9-35 illustrates this sequence of events, known as charge neutralization. We can show that the characteristic time for charge neutralization is $\tau_n = \epsilon_0/\sigma$. The conductivity of a gas used in this process is

$$\sigma = \frac{ne^2}{m\nu} \qquad (34)$$

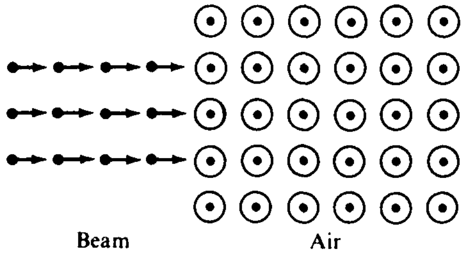
where n is the electron density, e and m the electron charge and mass, and ν the frequency of electron collisions of the heavy particles (ions or neutrals) that retard their flow. If we assume that the beam ionizes all standard temperature and pressure (STP) air, then $n \simeq 3 \times 10^{25}/m^3$ and $\nu \simeq 10^{14}/sec$ giving $\sigma = 8.43 \times 10^{3}$ mho/m and $\tau_n = 10^{-15}$ sec. This time is very short compared to the doubling time for beam size presented earlier. Thus, we can conclude that charge neutralization may prevent catastrophic beam spreading. The exact dynamics of the process are still the object of research, both theoretical and, to the extent possible, experimental.

What will happen to the beam after charge neutralization? The charge density within the beam will be zero, but the current density will not. This means that the repulsive E_r has been eliminated, but the attractive has not (see equation 32). Now the beam will feel a net inward force, and pinch down. At beam edge ω, this inward magnetic force is

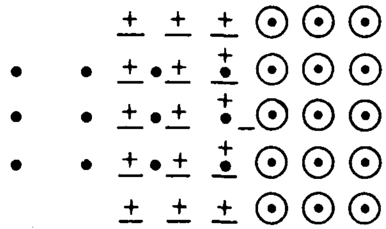
$$F_r = qvB_\theta = \frac{qv^2}{c^2}\left(\tfrac{1}{2}\,\frac{1}{\pi v\omega\epsilon_0}\right) \qquad (35)$$

What is the influence of this force on the size of the beam? Eventually, the small sideways motion, which relates to the temperature of the beam particles, will produce an outward pressure to counter the inward magnetic force, and the beam will stabilize as illustrated in figure 9-36.

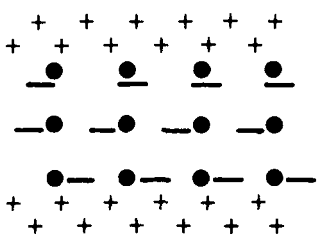
The preceding discussion shows that a CPB begins to expand under the repulsion of its charges as it enters the atmosphere, but it contracts after it ionizes the air and charge neutralization occurs. As the beam propagates, it will encounter particles of air. In scattering from the particles of air, the beam particles will pick up more sideways motion, and increase in the beam temperature will cause it to begin spreading again. In other words, as scattering heats up the beam, it will expand and expend energy in working against the compressive magnetic field. Thus,



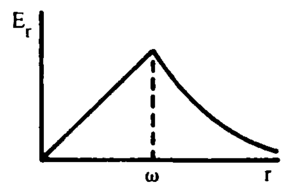
(a) Beam particles encounter air molecules as they exit the accelerator



(b) Beam particles strike the air molecules and ionize them, leaving a conducting plasma behind the head of the beam



(c) Charges of like sign are repelled from the beam volume, while those of opposite sign are attracted. The net result is that the charge density goes to zero, and a radial electric field which once looked like this,



now looks like this,

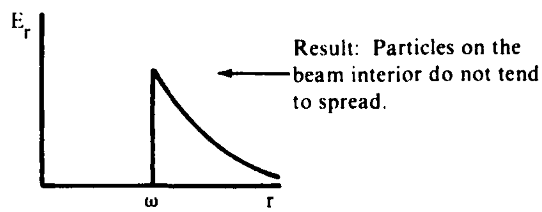
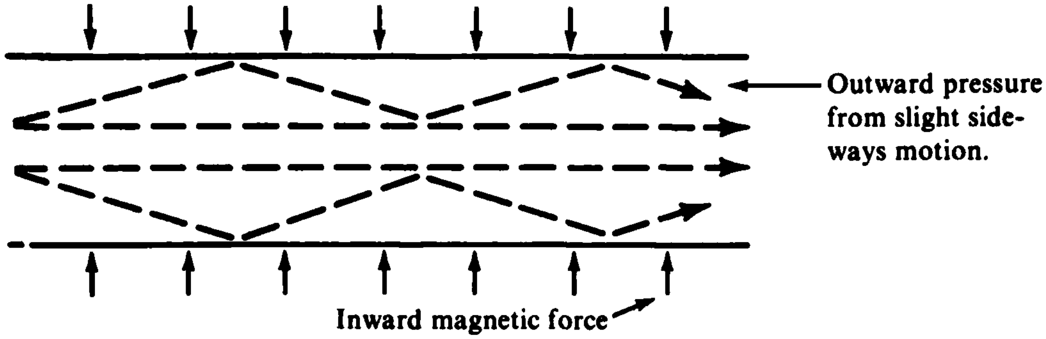


Figure 9-35. Charge neutralization.



$$F_r = \frac{qv^2}{c^2} \left(\frac{1}{2} \frac{I}{\pi v \omega t_o}\right)$$

Figure 9-36. Steady-state beam radius.

$$dQ = dW \tag{36}$$

where dQ is an increment of random energy (temperature) added to the beam through scattering and dW is an increment of work through beam expansion. A glance at each side of equation 36 shows that the work done is

$$dW = F\,d\omega \tag{37}$$

if the beam expands from a radius ω to $(\omega + d\omega)$ against a compressive force F, where the force felt per unit of length of beam is

$$F = \underbrace{\pi\omega^2 n_b}_{\text{particles per unit length}} \underbrace{\frac{qv^2}{c^2} \left(\frac{1}{2} \frac{I}{\pi v \omega \epsilon_o}\right)}_{\text{force per particle}} \tag{38}$$

It is important to note that the number of particles per unit length in the beam is a constant. As ω goes up, n_b goes down in such a way that $\pi\omega^2 n_b$ remains fixed. Similarly, the beam current is independent of beam size. Consequently, equations 37 and 38 indicate that

$$dW = C\frac{d\omega}{\omega} \tag{39}$$

where

$$C = \pi\omega^2 n_b\, q\frac{v^2}{c^2} \left(\frac{1}{2} \frac{I}{\pi v \epsilon_o}\right) \tag{40}$$

is a constant. What about the dQ term in equation 36? As the beam propagates, it will encounter air molecules in direct proportion to the distance it travels. Thus, we can conclude that

$$dQ = \left(\frac{dQ}{dz}\right) dz \tag{41}$$

where dQ/dz, the random energy of motion transferred to the beam per unit path length, is a constant. By substituting equations 39 and 40 into equation 36, we find

$$\left(\frac{dQ}{dz}\right) dz = C \ \frac{d\omega}{\omega} \quad . \tag{42}$$

We may solve equation 42 for the variation for ω with z to find

$$\omega = \omega_0 \, e^{\frac{1}{C}\left(\frac{dQ}{dz}\right) z} \quad . \tag{43}$$

We know equation 43 as the Nordseick Equation. It shows that the beam expands exponentially with distance, with the rate of expansion being directly proportional to (dQ/dz) and inversely proportional to I through the constant C as equation 40 shows. One consequence is that a minimum current is necessary to achieve a given range with little beam spread and that the minimum current required increases with range. This explains why large beam currents are required for long-range propagation. All of this is summarized conveniently in figure 9-37, which is a schematic representation of beam radius as a function of range.

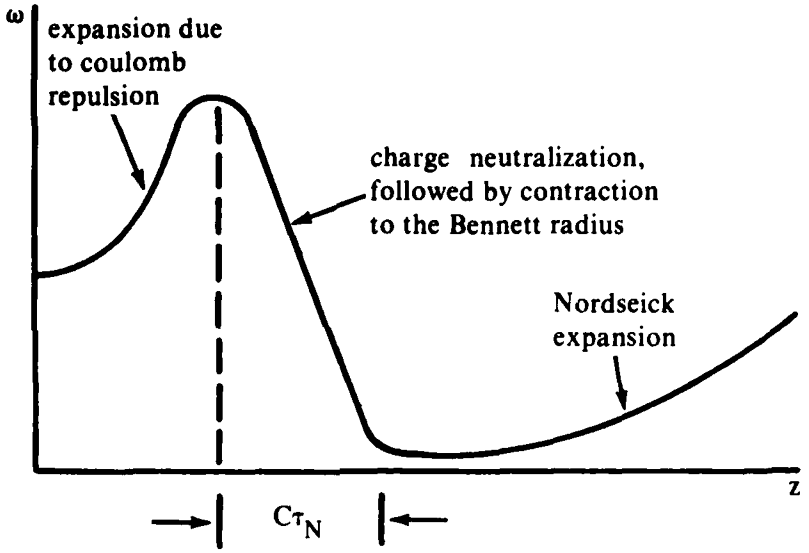


Figure 9-37. Beam size versus range (not to scale).

What if we choose the beam parameters so that little expansion will occur over the range desired for some application? Does this mean that the beam will propagate for long ranges? This is not true. As is true of lasers, beam expansion is only part of the story. Beam particles suffer loss of energy when they encounter air molecules. This process is analogous to the absorption of photons from a laser.

Energy loss in atmospheric encounters can be a key limitation to the usefulness of CPBs. When a particle of the beam encounters a molecule of air, it can lose energy by one of three processes: ionization, bremsstrahlung, or nuclear interactions. Ionization of air molecules by the beam particles forms the conducting channel responsible for charge neutralization and CPB propagation. Ionization has disadvantages as well as advantages, since every ionizing process requires that a beam particle must lose an amount of energy equal to the ionization potential, approximately 10-20 eV.

Since this energy is small in comparison to particle energies in the MeV-GeV range, a single beam particle can create many ion pairs before it loses all of its energy. This characteristic of particle beams is in contrast with the photons for a laser, which are totally absorbed in one interaction. One consequence is that a CPB is particularly effective against hard targets because it can penetrate more deeply into a target of solid density than a laser beam can.

Bremsstrahlung, a German word for "braking radiation," means that a charged particle emits radiation when it is accelerated or decelerated. Thus, if a particle in the beam comes close enough to an air molecule that the molecule deflects it, it will emit radiation and lose some energy. As with ionization, the amount of energy lost in bremsstrahlung is small compared to the total particle energy. The process requires many encounters to absorb all of the energy carried by a particle.

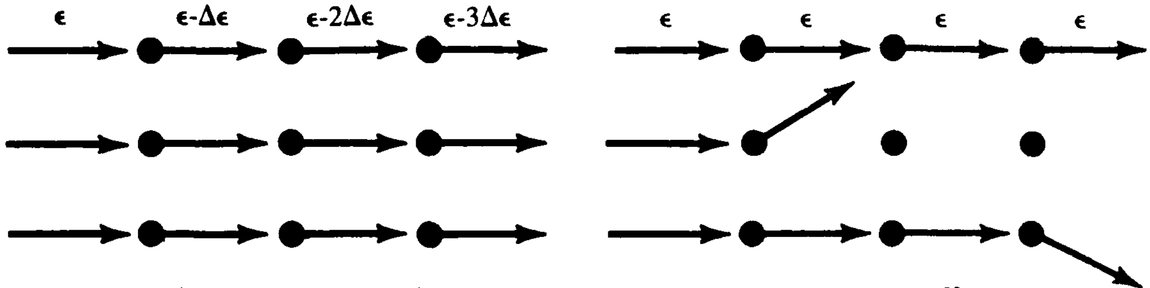
Nuclear interactions are somewhat different. You may recall from a modern physics course that the atomic nucleus occupies only a very small fraction of the atomic volume (approximately 10^{-12}). Therefore, collisions of beam particles with nuclei are relatively rare in comparison to collisions with the orbital electrons. However, when they do occur, they can be quite catastrophic, since the binding energies of the protons and neutrons comprising the nucleus are comparable to the energies of the beam particles. Therefore, the removal of a particle from the beam requires at most only a few nuclear interactions.

The effect of ionization and bremsstrahlung is to reduce the beam's energy and leave its current constant. The beam current is

$$I = n_b q v \pi \omega^2 \tag{44}$$

as long as $\gamma > 1$, $v \simeq c$, independent of the beam energy. As the beam particles move along, they lose energy, and their density and velocity remain constant, leaving I unaffected.

However, nuclear interactions remove particles from the beam and thus, reduce beam current. The energy per particle for those particles remaining in the beam is unaffected. Figure 9-38 illustrates the different ways by which ionization, nuclear interactions, and bremsstrahlung affect beam parameters.



(a) Ionization and Bremsstrahlung affect particle energy.

(b) Nuclear Events affect beam current.

Figure 9-38. Effect of interactions on beam parameters.

How do these various processes depend on beam and atmospheric parameters? We can show that $d\epsilon/dx$, therefore, we can write the energy loss rate per unit path length of travel due to ionization as

$$\frac{d\epsilon}{dx} = \frac{1}{4\pi\epsilon_0{}^2} \frac{NZq^2e^2}{mv^2} \left[\tfrac{1}{2} + \ln\left(b_{max}/b_{min}\right) \right] \tag{45}$$

where b is the distance between the beam particle and the atomic orbital electron, q is the charge of the particle, e is the electron charge, m is the electron mass, N is the number of atoms per unit volume, and z is the number of electrons per atom. Typically, $b_{max}/b_{min} \simeq 10^5$ and is proportional to γ. Therefore, a plot of $d\epsilon/dx$ versus $(\gamma - 1)$ should look like the sketch in figure 9-39.

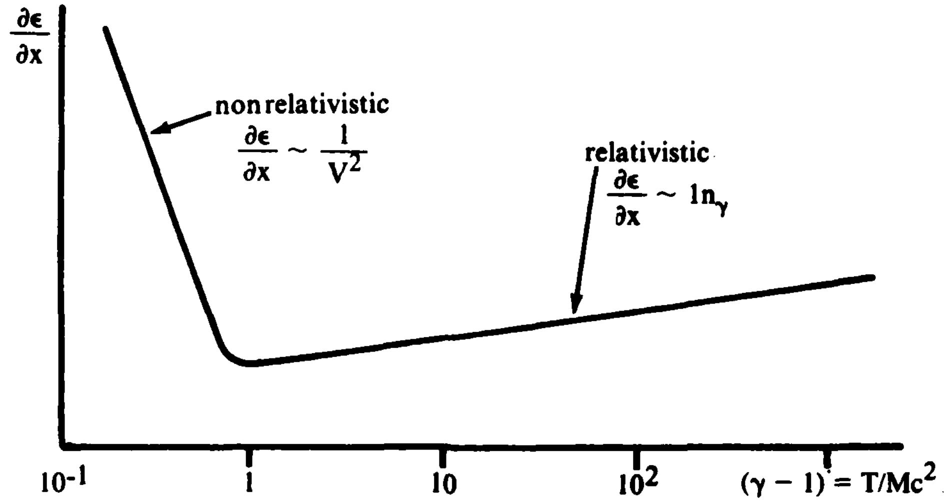


Figure 9-39. $\frac{\partial \epsilon}{\partial x}$ for ionization versus particle kinetic energy.

We can see from equation 45 and figure 9-39 why relativistic particles are of interest. For $(\gamma - 1) \gtrsim 1$, $d\epsilon/dx$ is minimized. If we take $\gamma \simeq 2$, $v \simeq c$, and $b_{max}/b_{min} \simeq 10^5$, we can estimate the minimum value of equation 45 as

$$\left.\frac{d\epsilon}{dx}\right|_{min} = \frac{1}{4\pi\epsilon_0^2} \frac{NZq^2e^2}{mc^2} \times 12 \tag{46}$$

For standard temperature and pressure (STP) air, $N = 3 \times 10^{19}$ and $Z \simeq 14$, giving $(d\epsilon/dx)_{min} = 4 \times 10^{-14}$ j/m $= 2.5 \times 10^5$ eV/m. Based on this computation, the range of a 1 MeV particle should be approximately 4 meters, and 4 kilometers for a 1 GeV particle. Of particular importance is the fact that equation 45 does not depend upon the mass of the beam particle. At the same value of γ, it predicts that electrons and protons will have the same range.

When a beam particle encounters an air electron in the bremsstrahlung process, it transfers a momentum $\Delta\rho = 2qe/4\pi v\epsilon_0 b$ to the electron. Since momentum is conserved, the beam particle must receive a kick in the opposite direction; something must accelerate the particle. Any text on electromagnetic theory will reveal that, when something accelerates a particle, the particle will radiate and when the acceleration is in a direction perpendicular to the direction of motion [$\dot{v} \perp v$], the power is radiated

$$P = \frac{2}{3}\frac{q^2}{4\pi\epsilon_0}\frac{\dot{v}^2}{c^3}\gamma^4 = \frac{2}{3}\frac{q^2}{4\pi\epsilon_0 c^3}\left(\frac{F}{M}\right)^2 \gamma^2 \tag{47}$$

How does this radiated power affect beam particle energy? We can estimate the total loss rate due to bremsstrahlung by

$$\frac{d\epsilon}{dx} = 2\pi NZ \int_{b_{min}}^{b_{max}} \Delta\epsilon(b) \; b\,db \simeq \frac{2\pi NZq^2}{4\pi\epsilon_0 c^3 m^2} \left(\frac{qe}{4\pi\epsilon_0}\right)^2 \frac{\gamma}{v b_{min}} \tag{48}$$

We should note that the $d\epsilon/dx$ is proportional to γ. Since $\epsilon = \gamma mc^2$ it follows that we can write equation 48 in the form

$$\frac{d\epsilon}{dx} = -\frac{\epsilon}{X_0} \tag{49}$$

where X_0 (a conglomeration of constants we can find from equation 48) is a radiation length. For STP air and $M = m_e$, $X_0 \simeq 300$ m.

Comparison of equations 49 and 45 shows that we may contrast $\partial\epsilon/\partial x$ from ionization and bremsstrahlung. We can show this in table 9-1. Therefore, bremsstrahlung is of greater importance for electrons at $\gamma \simeq 20$, and for protons at $\gamma \simeq 2 \times 10^7$. Thus, bremsstrahlung is not a practical consideration for other than electron beams. However, nuclear interactions are important only for proton and other heavy particle beams, since electrons cannot perturb the nucleus through the strong, short-range nuclear force. When a proton collides with a nucleus the result is a typical atom smashing. The nucleus will split and produce many species. In turn, these species may collide with other nuclei, and a cascade of the type shown in figure 9-40 will result. Detailed calculations of these events involve Monte Carlo analyses of the numerous possible interactions and their outcomes. However, in the absence of these analyses, we can draw a few general conclusions. A typical cross section for reactions of this sort in air is 230 millibarns (1 barn = 10^{-28} m) and is relatively constant over the range 0.1-10 GeV. The resulting proton mean free path is $l = \frac{1}{N\sigma}$, which for atmospheric density ($N = 3 \times 10^{19}/cm^3$) and $\sigma = 230$ mb gives $l = 1.4$ km. When we considered the effects of ionization, we saw that GeV protons would have a range of about 4 kilometers if limited by that mechanism. Thus, we can see that we must consider the nuclear effects in any analysis of heavy particle propagation.

Several other factors can affect the propagation of a particle beam. Even if we choose the beam current and other parameters to allow propagation to the target, there is always the possibility of an *instability*. Some small fluctuation in beam parameters might generate a restoring force whose magnitude is so great that it can cause an even greater fluctuation in the opposite direction. If this process repeats until it destroys all of the beam's integrity, the instability will occur. Many electrical circuits show a similar effect, for example, when feedback increases the magnitude of oscillations, and the oscillations grow without bound. From the standpoint of beam propagation,

Table 9-1
Ionization Versus Bremsstrahlung

Table 9-1
Ionization Versus Bremsstrahlung

	Energy Dependence ($\gamma \gtrsim 2$)	Particle Mass Dependence
Ionization	Weak, $\sim ln\epsilon$	Particle Independent
Bremsstrahlung	Strong, $\sim\epsilon$	$\sim 1/m^2$

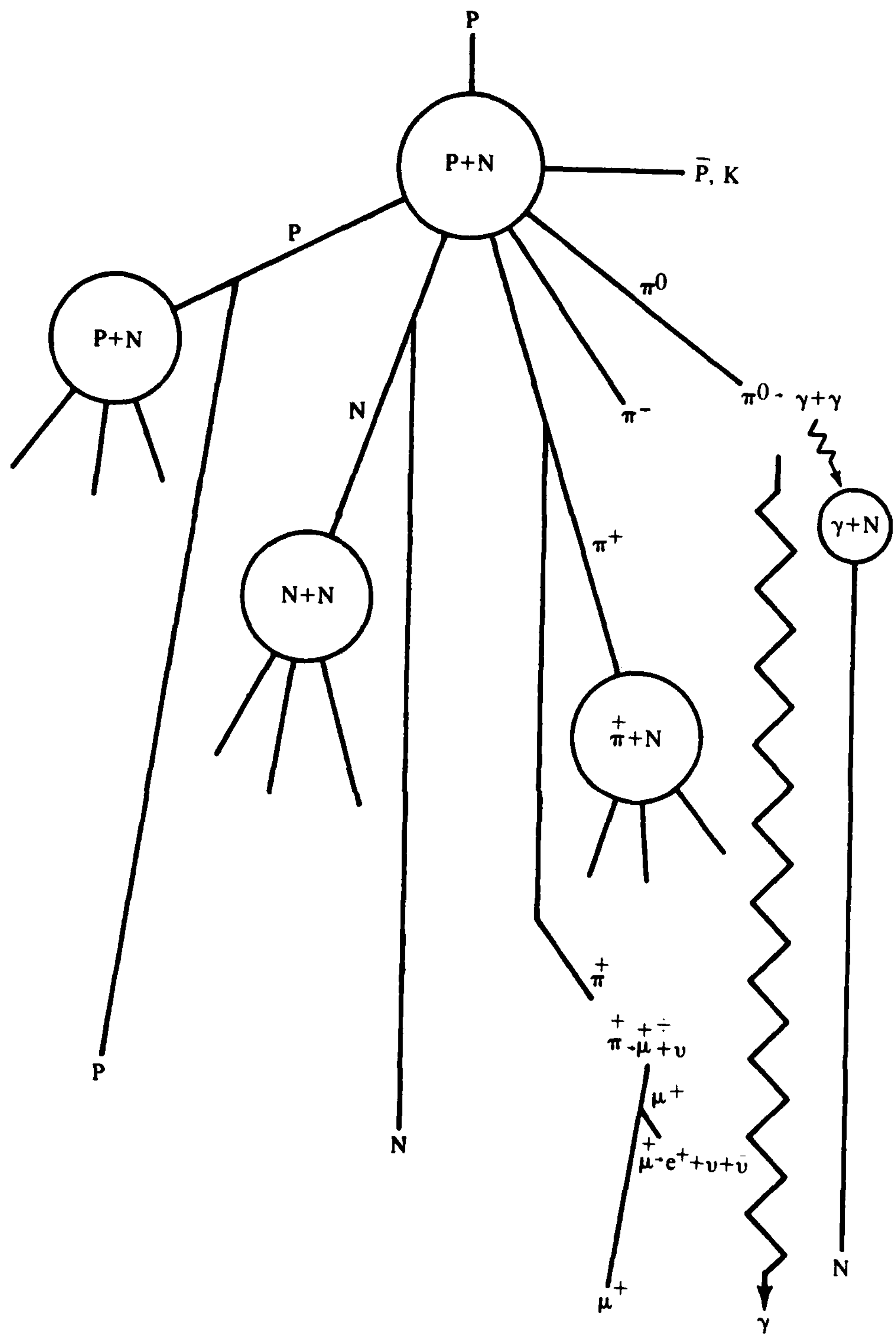


Figure 9-40. Intranuclear cascade.

the most important instability is the resistive hose instability, which results from the fact that the beam must propagate through an ionized conducting medium (the air that it ionized). As equation 27 shows, a magnetic field in the Θ-direction encircles the beam as it propagates (fig. 9-41).

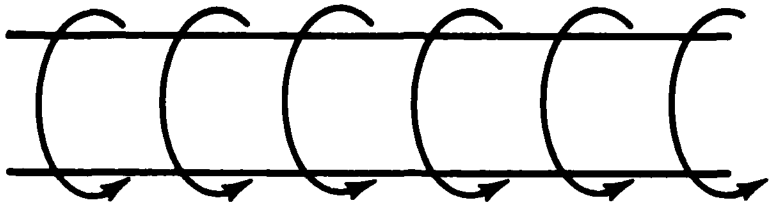


Figure 9-41. CPB and encircling magnetic field.

What will happen if the beam wanders from its original axis? Currents set up in the background air plasma will result in a first-order B-field that cancels out and adds to the B-field of the beam in such a way that the total magnetic field remains unchanged. This is because $\partial B/\partial t = -\text{curl}\ \underline{E}$. If the beam is charge-neutralized, there will be no $\underline{E}$ on its interior; and the total magnetic field must be constant. To a first approximation, the magnetic field lines are frozen into the background plasma. As a result the beam will not propagate in a stable manner, but will oscillate wildly the further down range it goes and finally will lose all of its integrity, as figure 9-42 illustrates. The only way around this problem is to "chop" the beam by pulsing it in segments too short and too widely separated to allow the feedback from one portion of the beam to another. The proper length to chop is something shorter than the e-folding length over which the perturbation grows.

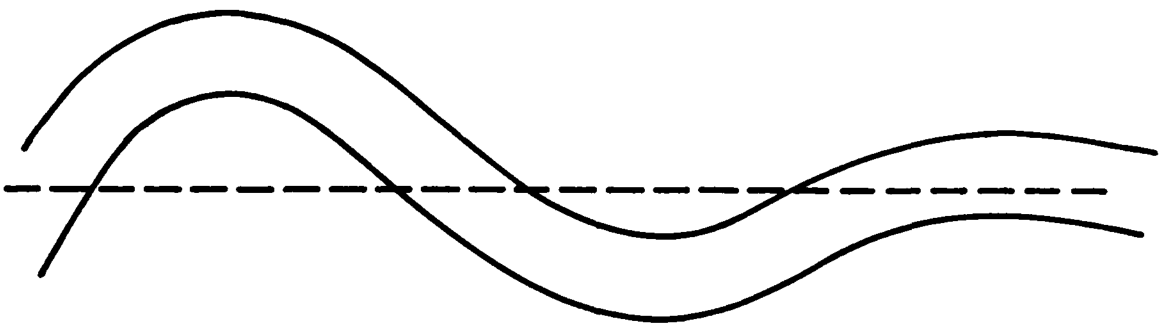


Figure 9-42. Beam behavior under influence of the hose instability.

Implications

CPBs are inherently more difficult than lasers to propagate over long distances. Two implications of this analysis are that space applications require neutral beams and that long-range propagation in air is a difficult process. Even a neutral beam must begin its life as a CPB, since we can accelerate only charged particles through electromagnetic forces. Thus, most schemes for making neutral beams begin with some electronegative atom, such as H, which can attach an electron, for example,

$$H + e \rightarrow H^- . \qquad (50)$$

The negative hydrogen ions are accelerated through one of the schemes discussed earlier. Following acceleration, the electrons are stripped collisionally from the H and, in effect, the process reverses

equation 50. We show schematically a setup to produce a neutral particle beam in figure 9-43. The major disadvantage of this arrangement from the standpoint of propagation is the stripping process, where collisions introduce divergence and spread into what would be a well-collimated beam otherwise.

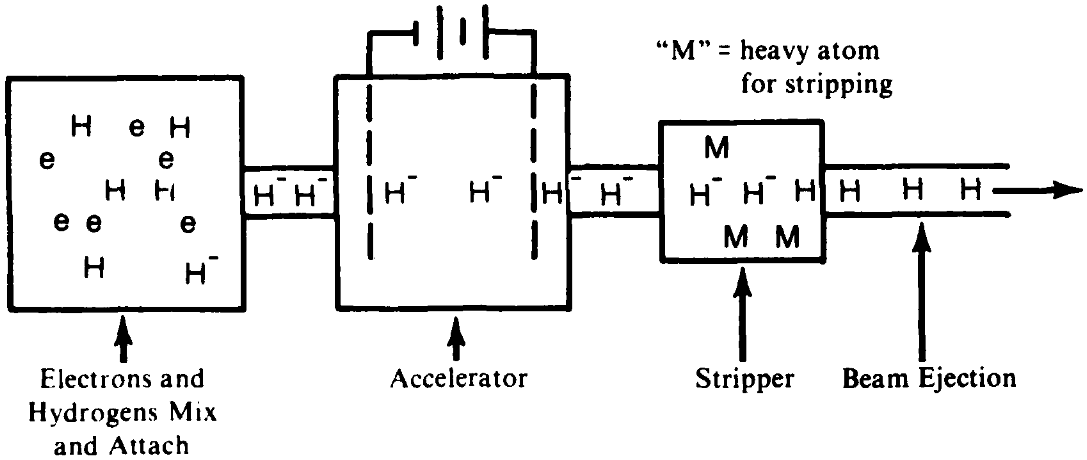


Figure 9-43. Schematic of a neutral particle beam generator.

For long-range CPB propagation in air, both ionization and nuclear interactions limit ranges to roughly a kilometer in STP air. Since such a range is unacceptable for most weapons applications, the designer must find a way to avoid this condition. One possible way is known as "hole-boring" (use of the beam itself to create a channel of less dense air through which it can better propagate). The beam heats the air through which it passes, and the energy loss expressed in equation 45 is an energy gain to the surrounding air. This air will expand and become less dense, much the same as the process in thermal blooming but to a much greater extent. Since the losses to all types of atmospheric encounters are proportional to air density, it follows that the range will increase as N decreases. In this way, we can achieve longer ranges, but only at the expense of potential problems from fast-moving targets, which may move out of the "hole" before being destroyed. Thus hole-boring poses problems for the designer of the accelerator, who must meet constraints of pulse width and separation to insure viability of the concept. As is true in thermal blooming, the time scale for opening and closing a hole is $\tau_h \simeq \frac{\omega}{a}$ where a is the speed of sound and ω the beam size. Opening a hole all the way to the target takes time, like τ_h, multiplied by as many STP beam ranges as there are between the accelerator and the target.

Summary

Table 9-2 shows similarities and differences between lasers and charged particle beams. In both instances, many small "bullets" fired at velocities at or near the speed of light accomplish the target damage. However, major differences in the absorption characteristics of these "bullets" have significant implications from the standpoint of damage criteria.

TARGET DAMAGE

What happens when a laser or a particle beam hits a target? This discussion treats CPBs and neutral beams together since both act as CPBs once they enter the targets. In science fiction, targets are vaporized completely; but destruction of targets in the real world is far less dramatic.

Table 9-2.
Lasers Versus Particle Beams

Table 9-2
Lasers Versus Particle Beams

	Lasers	Particle Beam Accelerators
Basic Elements of Design	An active medium A means of excitation An optical cavity Method for beam extraction	Source of particles Accelerating E-field Beam-handling techniques Particle injection/extraction
Some Basic Types	Electrical Chemical Gas-dynamic	Linear Circular Collective Effect
Type of "Bullets" Fired	Photons $V = c, \epsilon = h$ $\simeq 1$ eV	Protons or Electrons $v \simeq c, \epsilon = \gamma mc^2$ $\simeq 1$ GeV
Propagation Issues		
Absorption	Catastrophic, $\Delta\epsilon = h\nu$ Strongly ν-dependent (propagation windows)	Gradual, $\Delta\epsilon >> \gamma mc^2$ Weakly dependent on ϵ (no windows)
Spread	$w = w_0 [1 + (\frac{\lambda z}{\pi w_0^2})^2]^{1/2}$	$w = w_0\, e^{\frac{1}{c}(\frac{dQ}{dz})z}$
Interaction with Surrounding Air	Thermal Blooming	Charge Neutralization Hole-boring

Nevertheless, heating, melting, and vaporization are the key effects of energy deposited from the beam.

Basic Ideas

The heat capacity (c) of some material is a measure of the amount of energy required to raise its temperature by one degree. Therefore, c has dimensions of erg/gm K and joules/kg K in centimeter grams second (CGS) and meter kilogram second (MKS) units, respectively. Water has a heat capacity of 4.2 j/gm K, which means that it requires 4.2 joules of energy to raise 1 gram of water one degree in temperature. In contrast, aluminum has a heat capacity of 0.9 j/gm K. Obviously, water makes an effective coolant because it can absorb relatively large amounts of energy with little rise in temperature. A c_v measures at constant volume, and a c_p measures at constant pressure. Real-world situations are ones at constant pressure, but in any event, c_p and c_v at room temperature and above are close enough together that we can use either one.

The heat of fusion (L_m) is the amount of energy required to convert a given amount of some material from a solid at the melting point to a liquid at the melting point (erg/gm or joules/kg). Typical values are 334 j/gm for H_2O, and 1.1×10^4 j/gm for aluminum (Al).

Similarly, the heat of vaporization (L_v) is the amount of energy required to convert a given amount of some material from a liquid at the boiling point to a gas at the boiling point. Typical values are 2.44×10^3 j/gm for H_2O and 1.1×10^4 j/gm for aluminum.

We can use all three of these concepts to calculate the amount of energy required to boil away 1 gram of aluminum initially at room temperature. First, we must heat the metal from room temperature (300 K) to its melting point (933 K). This requires an energy

$$\epsilon_1 = mc(T_m - T_o) = (1\ gm)(9.0 \times 10^6 \frac{erg}{gmK})(933\ K - 300\ K) = 5.7 \times 10^9\ erg \quad . \tag{51}$$

Second, we must melt the solid aluminum to liquid aluminum. This requires an energy

$$\epsilon_2 = mL_m = (1\ gm)(4.0 \times 10^9 \frac{erg}{gm}) = 4.0 \times 10^9\ erg \quad . \tag{52}$$

Third, we must raise the molten aluminum at 933 K to its vaporization temperature at 2,333 K. This requires an energy

$$\epsilon_3 = mc\ (T_v - T_m) = (1\ gm)(9.0 \times 10^6 \frac{erg}{gm})(2333\ K - 933\ K) = 1.26 \times 10^{10}\ erg \quad . \tag{53}$$

Last, we must vaporize the molten aluminum, which requires an energy

$$\epsilon_4 = mL_v = (1\ gm)(1.1 \times 10^{11} \frac{erg}{gm}) = 1.1 \times 10^{11}\ erg \quad . \tag{54}$$

The total energy involved in vaporizing the aluminum is

$$\epsilon_T = \epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_4 = 1.32 \times 10^{11}\ erg \quad . \tag{55}$$

An interesting thing becomes apparent in looking at equations 51 through 55. Most of the energy required to vaporize a gram of aluminum is necessary for converting the liquid to vapor at T_v (equation 54), but the energy required to heat and melt the aluminum is very small in comparison. This is true for almost all materials and is a manifestation of the fact that breaking the bonds holding atoms together as a solid or liquid requires more energy than simply heating the atoms and making them wiggle faster. Table 9-3 shows values comparable to the values in equations 51 to 54 for several materials. Since the thermal properties of metals are quite similar, we can derive estimates of order-of-magnitude damage for almost any target regardless of its composition.

Table 9-3
Thermal Properties of Various Metals

MATERIAL	c $(\frac{erg}{gmK})$	cT_m $(\frac{erg}{gm})$	L_m $(\frac{erg}{gm})$	cT_v $(\frac{erg}{gm})$	L_v $(\frac{erg}{gm})$
Al	9.0×10^6	8.4×10^9	4.0×10^9	7×10^{10}	1.1×10^{11}
Mg	1.0×10^7	9.2×10^9	3.7×10^9	1.4×10^{10}	5.3×10^{10}
Stainless Steel	4.6×10^6	9.2×10^9	2.5×10^9	1.6×10^{10}	6.2×10^{10}

The thermal conductivity (k_c) of some material is a measure of its capacity to conduct heat. It is an unstable situation when one part of a substance is hot and another part is cold. Consequently, energy will flow in proportion to any existing temperature gradient. The flux of energy (W/m^2) resulting from a temperature gradient gradT is

$$\underline{u} = -k_c \mathrm{grad}T \tag{56}$$

where k_c is the thermal conductivity.

$$\frac{\partial \epsilon}{\partial t} = - \operatorname{div} \underline{u} = k_c \nabla^2 T \tag{57}$$

gives the increase or decrease in energy density (w/m^3) from this flow of energy in some region of space. Since

$$\begin{array}{cccc} \Delta\epsilon = & \rho & c & \Delta T \\ (\frac{w}{m^3}) & (\frac{kg}{m^3}) & (\frac{J}{k_g k}) & (k) \end{array} \tag{58}$$

relates the energy density to the temperature through the heat capacity, we can convert equation 57 to a differential equation for the temperature distribution in some substance. This equation is

$$\frac{\partial T}{\partial t} = \frac{k_c}{\rho C} \nabla^2 T = D \nabla^2 T \tag{59}$$

where $D(m^2/sec)$ is known as the thermal diffusivity. The elements, k_c, ρ, and C can vary quite a bit from one material to another, but D is generally between $(1\text{-}10) \times 10^{-4} m^2/sec$ for most substances. Equation 59 is the heart of any target damage analysis that involves examination of the place where the weapon deposits energy, the method by which it deposits the energy, the effect of thermal conduction, and temperatures required to inflict damage.

Hard or soft target damage depends on the definition of damage and the temperature necessary to achieve that damage. For example, we consider a satellite blinded by a very small increase in temperature as soft. On the other hand, we consider reentry vehicles designed to withstand the high temperatures associated with reentry into the earth's atmosphere as very hard.

We have limited this analysis to hard targets in developing damage criteria for several reasons. First, the estimates are very conservative, since a laser or CPB with sufficient energy to damage a hard target will certainly have enough energy to damage a soft one. Second, analysis of hard targets produces very general results. Since damage to hard targets basically involves melting or vaporizing and is relatively insensitive to the type of material (see table 9-3), we can make reasonable estimates without specifying the target in much detail. However, soft targets are soft in different ways, and it would require much time and space to specify the nature of the target and the engagement scenario. Thus, a bonus of this approach is that the discussion involves unclassified material. Finally, we can use analysis of this type to establish damage thresholds for any target.

The 10 Kilojoule Criterion

We can draw an amazing conclusion from the previous discussion; 10^4 joules, properly applied, will destroy almost anything. We base this conclusion upon two premises. First, vaporizing 1 gram of almost anything requires approximately 10^4 joules (see table 9-3); second, removal of 1 gram of material from some vital spot will destroy almost any target. The second premise is the most difficult to accept. Practically all solid matter has a density of approximately 1 gm/cm^3; thus, we are eliminating only a very small amount ($\sim$1 cm^3) or breaking only a few of the bonds that hold the atoms of some substance together. However, removal of 1 cm^3 from an appropriate spot on almost anything will destroy it. For example, we can destroy an airplane by igniting its fuel, which requires putting a hole through the aircraft skin and the fuel tank. The total thickness of surfaces to be penetrated is undoubtedly less than 1 cm requiring energy (on target) of less than 10^4 joules. When a bullet penetrates a target, it pushes the target material aside, and breaks bonds only along a thin line of volume much less than 1 cm^3. Many more examples are possible, but the

point is that we can scrape up and weigh the remains of any target and never find more than a gram or so missing.

If the preceding arguments are not convincing, we might consider the fact that weapons systems throughout history have delivered energy on the order of 10^4 joules. For example, an ordinary rifle round has a muzzle velocity of 3,000 ft/sec or 9.1×10^4 cm/sec and mass of 180 grains, or 11.8 grams. Such a round has kinetic energy of

$$T = {}^1/_2\, mv^2 = 5 \times 10^3 \text{ joules} \tag{60}$$

within a factor of 2 to the 10^4 joules mentioned in this discussion. A more ancient example is the Roman siege catapult. According to a recent article in *Scientific American,* a typical catapult fired a stone with a mass of approximately 20 kilograms over a range of some 200 meters. If we assume that the firing of the stone at an angle of 45 degrees to the vertical of the range, we can calculate the kinetic energy of such a stone as

$$T = 4 \times 10^4 \text{ joules} \tag{61}$$

within an order of magnitude of our damage criterion. A recent article in the *Journal of Applied Physics* provides a more futuristic example. R. F. Benjamin and G. T. Schappert report that they used a laser pulse containing about 40 μj of energy to create a hole with a diameter of approximately 50 μm in a glass wall 1 μm thick. The volume of the hole was approximately 2×10^{-8} cm^3. The energy required to vaporize 1 cm^3 of glass with such a laser would be approximately 2×10^3 joules. Once more, we can see that the typical value of 10^4 joules is not too unreasonable. The obvious conclusion from this is that a viable weapon against hard targets must deliver 10^4 joules as a minimum to a target volume of 1 cm^3 in a time too short for the target to get rid of the energy. Any serious weapon must fire much more than 10^4 joules to compensate for losses suffered on the way to the target, and it must fire a sufficient number of times to compensate for the probability that it will not strike a vital spot on the first shot.

Damage Thresholds

We have shown in the discussion thus far the amount of energy required to inflict damage. It has provided a transport equation (equation 59) to determine the path taken by the energy from the point where it is deposited. We can combine these two elements to determine the thresholds at which lasers and charged particle beams will damage targets. A damage threshold involves the three factors of energy, area, and time. For example, a bomb releases a fixed amount of energy over a time of limited duration. However, as the burst of energy expands, it spreads over a larger and larger area, so that the pressure (force/area) felt by a target decreases with distance from the bomb, and, at some range, the target is safe. A laser beam firing over a shorter range than the Rayleigh Range has a constant cross-sectional area. However, the energy in the beam will decrease with distance because of atmospheric absorption, and may fall below the level of energy necessary for damage (especially if a cloud blocks the way). The sun delivers 5,000 j/cm^2 to the earth's surface in a 12-hour period, and this amount of energy is close enough to the 10 kilojoule criterion to produce some dramatic effects. What prevents it from killing people? Too much time is involved. Inert objects can get rid of the energy through radiation or thermal conduction, and living organisms have time to sweat, move into the shade, or use other cooling strategies.

Since energy, area, and time enter into all damage thresholds, usually the threshold is expressed in terms of beam intensity (W/cm^2), which accounts for all three factors. In keeping with this previous analysis, the energy absorbed per unit area should be something like 10^4 j/cm^2. An appropriate time scale depends on whether lasers or particle beams are involved because they deposit their energy in different ways. To understand why this is true, we must examine the interaction of lasers and particle beams with solid targets.

Interaction of lasers with solids. A substance absorbs laser light when the photon energy of the laser connects two energy levels of the substance. Thus, analysis of what happens to the energy levels of atoms and molecules when they come together to form solids should clarify the interaction of laser light with solids. As a general rule, aggregations of atoms have more degrees of freedom than individual atoms. For example, a single atom has only electronic degrees of freedom (see fig. 9-44), but a diatomic molecule has electronic, vibrational, and rotational degrees of freedom (see fig. 9-45). Thus, we would expect that a solid, much like a macromolecule of $\simeq 10^{26}$ atoms, should have many degrees of freedom and many corresponding energy levels. It should be much more likely to absorb a photon. The conclusion that a photon is lucky to be absorbed in a gas, and lucky to be transmitted in a solid, is a conclusion that reflects experience and intuition.

When two atoms are brought together to form a molecule, they perturb each other so that their individual energy levels split and become a group somewhere near the level of the original individual atoms (see fig. 9-46). When numerous atoms are brought together, there are so many possibilities that the individual lines blend together into bands separated by energy gaps. Within the band, all energies correspond to allowed states, and, in the gaps, there are no allowed states (see fig. 9-47).

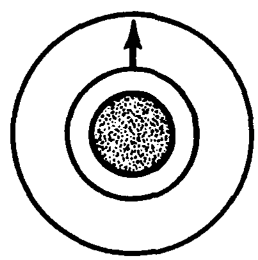


Figure 9-44. Degrees of freedom in a single atom.

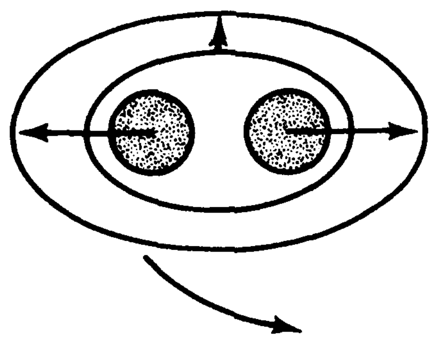


Figure 9-45 Degrees of freedom in a diatomic molecule.

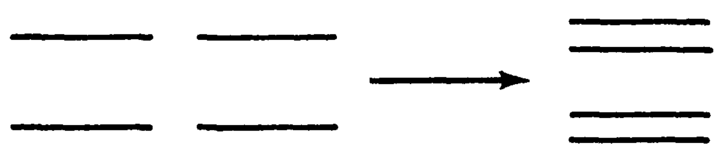


Figure 9-46 Atomic and molecular energy levels.

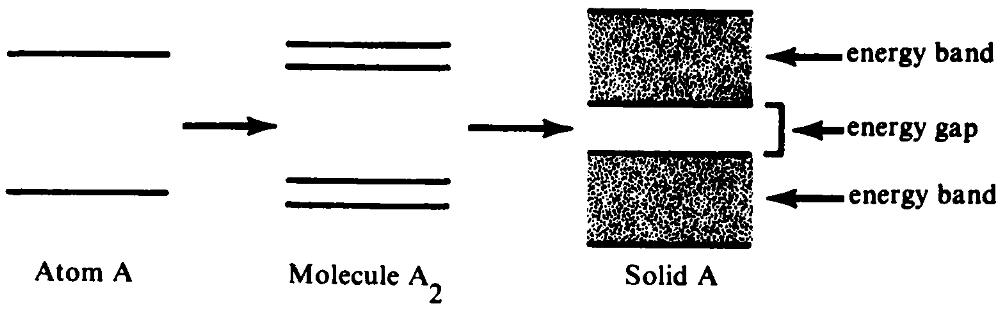


Figure 9-47. Energy band formation.

Broadly speaking, solids fall into two categories: metals and insulators. In an insulator, all the energy bands are either completely full or completely empty (of electrons). See figure 9-48. This means that electrons cannot be given energy in the lower band unless they are given at least ϵg, the energy required to get them into the upper band. Thus, DC E-field cannot accelerate the electrons in an insulator, since in time t it tries to raise an electron an amount $\Delta\epsilon = \int \rho E(dt/v)$, which lies in the energy gap or on another filled state. Therefore, insulators do not conduct electricity. Also, it is clear from this picture that insulators will be transparent to photons of frequency $\hbar\omega < \epsilon_g$ and opaque to those for which $\hbar\omega > \epsilon_g$. In the former case, empty and filled states cannot be connected, but they can be connected in the latter case.

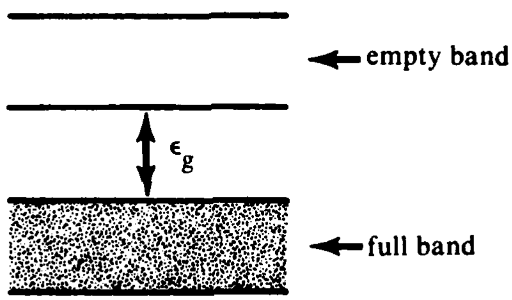


Figure 9-48. Energy levels in an insulator.

In a conductor, one band is partially filled. This means that the electrons in this band are free to be accelerated, and they conduct electricity (fig. 9-49). It means that light (an electromagnetic wave) will have a difficult time penetrating a conductor, since the electrons, which are mobile and free to flow, tend to move in such a way that they cancel the E and B fields associated with the wave on the interior. Because the light cannot penetrate and must go somewhere, it is reflected. This explains why metals make good mirrors and appear shiny.

We can show from Maxwell's equations that light will not penetrate a conductor or a plasma unless the frequency of the light exceeds the plasma frequency given by

$$\omega^2 p = \frac{ne^2}{m\epsilon_0} \tag{62}$$

where n, e, and m are the free electron density, charge, and mass, respectively. At solid density, $n = 10^{30}/M^3$, so that

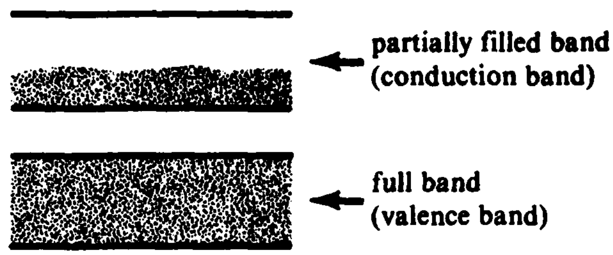


Figure 9-49. Energy levels in a conductor.

$$\omega_p \simeq \left(\frac{10^{30}/m^3(1.6 \times 10^{-19}\ \text{coul})^2}{9.11 \times 10^{-31}\text{kg}(8.85 \times 10^{-12}\text{fd/m})}\right)^{1/2} = 5.6 \times 10^{16}/\text{sec}.$$

This corresponds to a wavelength $\lambda = 2\pi c/\omega p \simeq 3 \times 10^{-8}$ m. Since visible light is in the 0.4 to 0.7 μm range, we can conclude that metals are highly reflective in the visible and, below that, they become transparent in the far ultraviolet to X-ray part of the electromagnetic spectrum. This reflects everyday experience.

At frequencies below ωp, we can use Maxwell's equations to show that light will decay going into a conductor according to $S = S_o e^{-x/\delta}$ where $S = (\frac{2c^2\epsilon_o}{\sigma\omega})$ is known as the skin depth. In this expression, σ is the conductivity of the material and ω the light frequency. A typical value for the skin depth, for $\sigma = 3.6 \times 10^7$ mho/m (aluminum) and 10.6 μm CO_2 radiation, is 1.6×10^{-6} centimeters.

The fraction of light reflected from the metal is related to the skin depth and is given by

$$R = 1 - \frac{2\omega}{C}\delta = 1 - 2\left(\frac{2\epsilon_o\omega}{\sigma}\right)^{1/2} \quad . \tag{63}$$

Reflectivity goes to 1 as the conductivity becomes large and as the frequency becomes low. This seems reasonable because deep penetration provides more opportunity for absorption of light. If ω is low, it is easier for the electrons in the metal to respond to the applied E and M fields in such a way as to exclude them. For aluminum and 10.6 μm radiation, substitution into equation 63 gives

$$R = 1 - 1.88 \times 10^{-2} = 0.981\,. \tag{64}$$

In other words, an aluminum target reflects 98 percent of the CO_2 laser light and only absorbs 2 percent. Obviously, such things as surface coating (paint), degree of polish, and other things strongly affect laser coupling to targets.

We have summarized the interaction of laser light with solids in table 9-4. According to the table, the absorption of laser light will be on the surface (and not in the interior) of a target if it takes place at all. Therefore, it is appropriate to model laser interaction with solids as a surface deposition of energy at an appropriate rate.

Interaction of charged particle beams with solids. The interaction of lasers with gases is rather simple in the sense that a molecular absorption spectrum characterizes it. Interaction with solids varies in complexity with differences between metals and insulators, and reflections and absorption.

However, the interaction of CPBs with gases is rather complex in the sense that we must consider nuclear interactions, ionization, and bremsstrahlung. Fortunately, time spent in contemplating these factors is worthwhile, since the interaction of CPBs with solids is exactly the same as their

Table 9-4.
Light Interaction with Metals and Insulators

Table 9-4
Light Interaction with Metals and Insulators

	Metals	Insulators
Transparent	$\omega > \omega_p$ (UV)	$\hbar\omega < E_g$ [$K \simeq 1\ km^{-1}$]
Absorbing	$\omega < \omega_p$ [~2% abs]	$\hbar\omega > E_g$ [$K \simeq 10^{9}\ km^{-1}$]
Reflecting	$\omega < \omega_p$ [$R \simeq 98\%$]	-----

interaction with gases—only the density is different. At energies of order 1 GeV, a particle beam has far more energy than any of the binding energies that distinguish a solid from a gas. Thus, to a CPB, a solid is much the same as a thick gas.

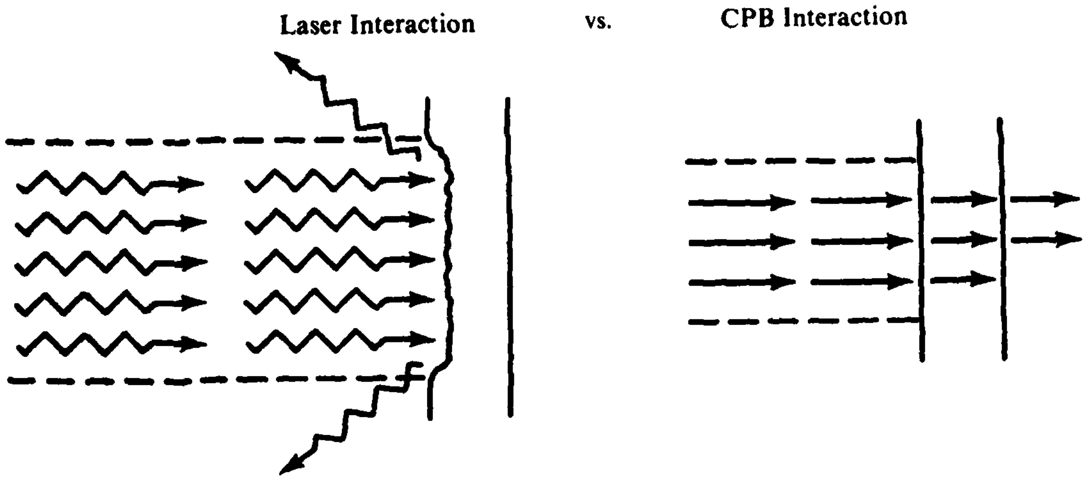
We have shown earlier that a GeV particle will have a range of approximately 1 kilometer if ionization is the mechanism for major energy loss. If the particle is an electron, bremsstrahlung will be the major energy loss and the range will be approximately 300 meters. Both of these results scale as 1/N, where N is the gas density. Thus, since solid matter is approximately 10^4 greater than that of STP air, we can estimate the range of protons in solids as approximately 10 centimeters, and the ranges of electrons approximately 3 centimeters. Since both of these estimates are greater than the thickness of a typical target (an aircraft skin, for example), two conclusions are obvious. First, CPBs, in contrast to lasers, will deposit their energy throughout the thickness of the target rather than on the surface. Second, propagating right through the target might waste a significant amount of the energy carried by a CPB. This wasted energy may be an advantage. In many cases the first target encountered (perhaps an aircraft skin) might not be the ultimate goal (perhaps a fuel tank). Thus, passing through the first target and proceeding to the second could be a definite advantage. Figure 9-50 compares the interaction of lasers and CPBs with solids.

Now we can establish the damage threshold for lasers since we know where they deposit their energy (on the surface), how the energy is transported (by thermal conduction), and what energy density is needed for damage, ($\sim$10KJ/cm^3). We can use the heat diffusion equation (equation 59) to predict the threshold laser intensity necessary to melt a hole in a target. If the location of interest is at a distance x_0 into the target, the laser will have transported heat to this portion of the target some time after it has engaged the surface (see fig. 9-51). A plot of temperature versus time at x_0 would appear as shown in figure 9-52. Similarly, a plot of temperature versus position at time T would appear as shown in figure 9-53. A glance at these two figures suggests that we may estimate the partial derivatives appearing in equation 59 as

$$\left.\frac{\partial T}{\partial t}\right|_{x_0} \simeq \frac{\Delta T}{\tau} \tag{65}$$

and

$$\left.\frac{\partial^2 T}{\partial x^2}\right|_{\tau} \simeq \frac{\Delta T}{x_0^{\,2}} \tag{66}$$



1) Energy deposition is a "front surface" effect — finite "burn through" time
2) Energy is wasted by reflection
3) Target can be made twice as hard by doubling its thickness

1) Energy deposition "in depth" — rapid engagement of critical components
2) Energy is wasted by transmission
3) Hardening is generally not practical

Figure 9-50. Laser interaction versus CPB interaction.

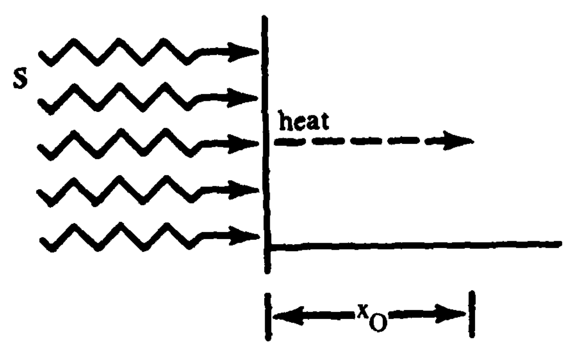


Figure 9-51. Heat flow.

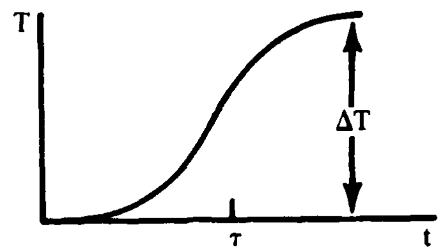


Figure 9-52 Temperature versus time at a given position.

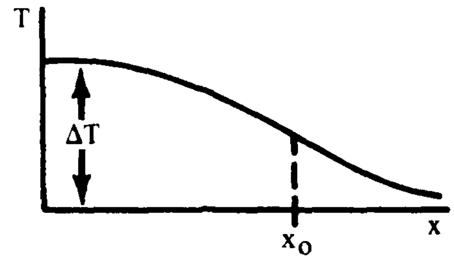


Figure 9-53 Temperature versus position at a given time.

Substituting these estimates into equation 59, we find

$$\frac{\Delta T}{\tau} = D\frac{\Delta T}{x_o^2} \quad . \tag{67}$$

We can solve equation 67 for the time, τ, necessary for heat to diffuse a distance x_o,

$$\tau = x_o^2/D \tag{68}$$

or the distance, x_o, to which heat will penetrate in time, τ,

$$x_o = (D\tau)^{1/2} \quad . \tag{69}$$

Equation 69 shows that, in time, T, a laser beam of radius w, will heat a volume

$$V = \pi w^2 x_o = \pi w^2 (D\tau)^{1/2} \quad . \tag{70}$$

During this time, the beam will have deposited an energy

$$E = \pi w^2 S\tau \tag{71}$$

where S is the beam intensity (W/m^2).

Consequently, the heated region will have an energy density

$$E/V = S\tau/(D\tau)^{1/2} = S(\tau/D)^{1/2} \quad . \tag{72}$$

The energy density necessary to reach the melting point is

$$E/V = \rho C(T_m - T_o) \tag{73}$$

where ρ is the mass density of the target (gm/cm^3), C the heat capacity (J/gm K), and $T_m - T_o$) the rise in temperature necessary to reach the melting point. Equations 72 and 73 show that T_m will be reached if $St^{1/2}/D = \rho C(T_m - T_o)$ or if

$$S = (D/\tau)^{1/2} \rho C(T_m - T_o) \quad . \tag{74}$$

Equation 74 indicates that the minimum intensity necessary to melt a target will occur for the maximum possible T; that is, when $T = t_p$, the pulse width of the laser. Thus, the *threshold intensity* for melting a surface of diffusivity, D, and heat capacity, C, and density p by a laser of duration T_p is

$$S_m = (D/t_p)^{1/2} \rho C T_m \tag{75}$$

where generally $T_m >> T_o$.

We must make a number of points regarding equation 75. First, it is based on the assumption that heat flows in one dimension, that is, the heat flows into the target rather than in a radial direction. On an intuitive basis, this assumption will be valid as long as $X_0 << W$ the beam radius.

Thus, a criterion for the validity of equation 75 is (refer to equation 69)

$$(Dt_p)^{1/2} < w \tag{76}$$

For w = 10 cm and D = 1 cm^2/sec, equation 76 requires $t_p < 100$ sec. Then equation 75 will be valid for beams of military significance, but it may be suspect for present-day beams with very small spots. Second, S_m is the intensity that must be absorbed if the target has a reflectivity, R. The actual beam intensity necessary to melt the target would then be $S = S_m/(1 - R)$. Since R is approximately 97 percent (see equation 64), this is an important point to remember. Finally, melting may or may not be sufficient for penetration of a target surface since penetration requires the removal of the molten material. In the best case, it will be flushed out by the flow across the surface. In the worst case, we must cause molten material to be heated further and vaporized for removal.

In the worst case, we can use an analysis exactly analogous to the analysis for the melting threshold to find the threshold of vaporization. Equating the energy deposited with the energy required to reach T_v,

$$S_V \left(\frac{t_p}{D}\right)^{1/2} = \rho C(T_m - T_o) + \rho L_m + \rho C(T_V - T_m) \simeq \rho C T_V \tag{77}$$

or

$$S_V = (D/t_p)^{1/2} \rho C\, T_V \quad . \tag{78}$$

The caveats that apply to equation 75 apply to equation 78. Since R may be much less in the molten state than in the solid state, the actual laser intensity necessary to reach S_V may not be much greater than the intensity necessary to reach S_m. Figure 9-54 shows these intensities for aluminum.

If S exceeds the threshold for vaporization, how fast can the laser beam drill through the target? The simple use of energy balance can provide the answer to the question. For example, if the surface is being eroded (the hole is getting deeper) at a velocity V_d, the material being removed has been raised to an energy density

$$\rho C(T_m - T_o) + \rho L_m + \rho C(T_V - T_m) + \rho L_V \simeq \rho L_V \quad . \tag{79}$$

Thus, energy flows away from the surface at a rate $\rho L_v V_d$(W/m^3). Under conditions of steady-state erosion, the energy absorbed from the laser must balance this loss of energy. That is expressed as

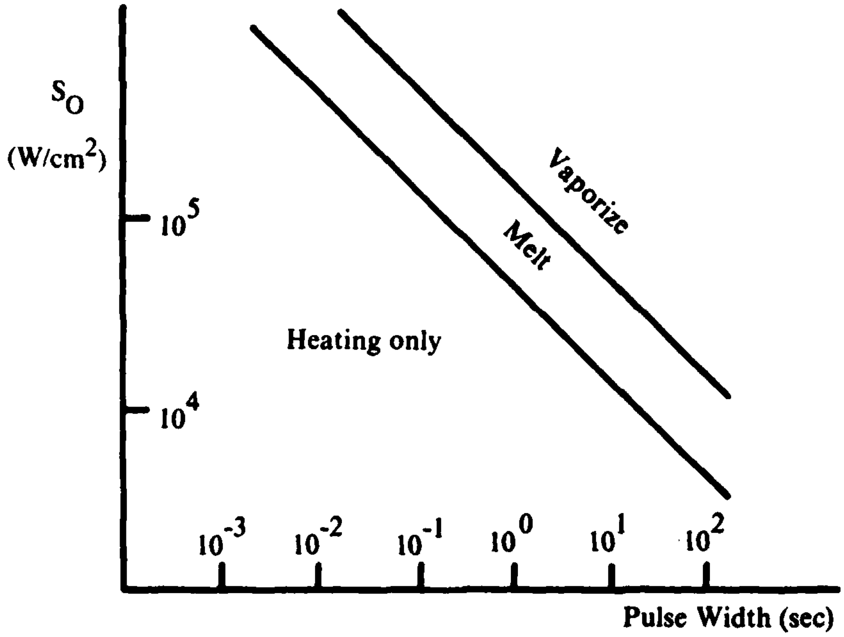


Figure 9-54. Melting and vaporization thresholds for aluminum.

$$S = \rho L_v v_d$$

or

$$v_d = S/\rho L_v \quad . \tag{80}$$

Similarly, if we need only to melt the surface to drill through, the rate of erosion would be

$$v_d = S/\rho(CT_m + L_m) \quad . \tag{81}$$

What are some typical values? Recall that $\rho L_v = 10^4$ joules/cm^3. Thus, for $S = 10^5$ W/cm^2, equation 80 predicts $V_d = 10$ cm/sec. In other words, a beam of this intensity could burn through a 1 centimeter target in 0.1 seconds. During this time, the beam would have deposited 10^4 j/cm^2 of energy.

Usually, heating, melting, and vaporizing with subsequent hole-drilling are classed as thermal effects since heat deposited by the laser on the target caused them. However, once vaporization begins, the reaction force from the evolving vapor might put a high pressure over the surface and contribute to such mechanical effects as buckling and spalling, among others. This is done by driving shocks through the solid target. It is possible that this condition might alleviate these rather long burn-through times.

Pressure in a gas relates to temperature and density of gas molecules, n, by

$$p = nkt \tag{82}$$

In this case, we have vapor of mass density, ρ_v, evolving from the target surface at temperature T_v. Thus, $n = \rho_v/M$, where M is the mass of a vapor molecule. Thus,

$$p = \frac{\rho_V}{M} kT_V \tag{83}$$

Equation 83 shows that we must estimate ρ_V in order to estimate p. From conservation of mass, it follows that the density and velocity of the vapor are related to the density and erosion rate of the surface by

$$\rho_V v_V = \rho v_d \quad . \tag{84}$$

Moreover, since the temperature T_v at which the vapor evolves is translated into its macroscopic flow, we expect

$$v_V \simeq (kT_V/M)^{1/2} \quad . \tag{85}$$

We may substitute equation 85 into equation 84 to give ρ_v in terms of kT_v and V_d. We may substitute this result, together with equation 80, into equation 83 to yield

$$p = \frac{S}{L_V} \left(\frac{kT_V}{M} \right)^{1/2} \tag{86}$$

By recalling that $L_V = 10^4$ joules/gm, and that $kT_V/M \simeq CT_V \simeq 10^3$ joules/gm, we can conclude that equation 86 predicts $p \simeq 1$ joule/cm^3 for a laser of absorbed intensity 10^6 W/cm^2. By recalling that atmospheric pressure is approximately 0.1 joule/cm^3, we can conclude that an overpressure of 10 atmospheres is predicted in this case. Quite possibly, this may inflict some damage.

Once it becomes energetically possible to put pressure over a target, we can ask logically what the threshold is for inflicting a certain amount of damage or creating a desired level of stress over the surface. One answer is to plot pressure versus impulse data for desired damage levels. The result would appear as a hyperbola as figure 9-55 illustrates.

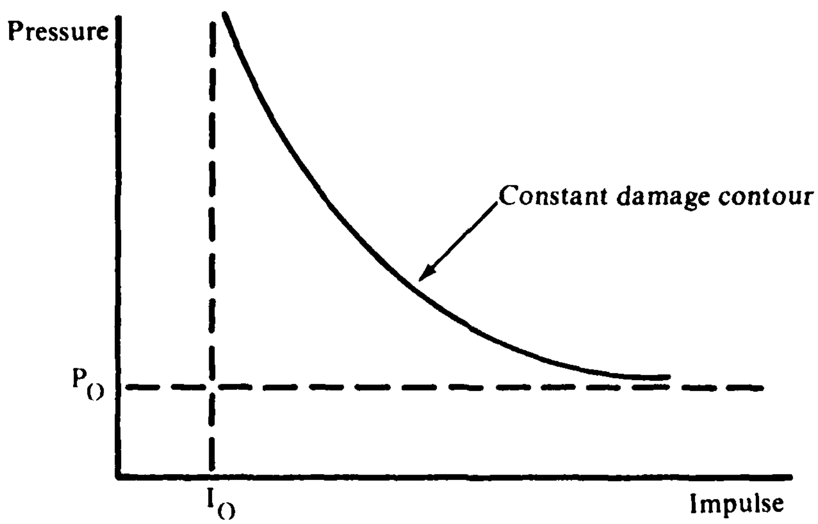


Figure 9-55 Isodamage contour

Impulse is the integral of force over time or approximately

$$I = p\pi w^2 t_p \quad . \tag{87}$$

Therefore, the data obtained would take the form

$$(p - p_o)(pt - I_o/\pi w^2) = D \tag{88}$$

where the constants p_o, I_o, and D would depend on the type of target as well as the desired level of damage. Equation 88 is in accord with earlier discussions. We can trade off pressure and application time, but we must exceed certain thresholds for each to inflict the desired damage.

What are the implications regarding damage thresholds? For pressures far in excess of P_o, equation 88 is approximately

$$p^2 t = D \tag{89}$$

However, if a laser of intensity S vaporizes a surface, we know the pressure (see equation 86). Substituting equation 86 into equation 89, we get

$$\frac{S^2}{L_V^2}\left(\frac{kT_V}{M}\right)t_p = D \quad . \tag{90}$$

Thus, the intensity necessary to achieve the level of damage specified by the constant D is

$$S_D = L\left(\frac{DM}{kT_v t_p}\right)^{1/2} \quad . \tag{91}$$

Interestingly, S_D is proportional to $1/\sqrt{t_p}$, just as S_m and S_r were (see equations 75 and 78). Figure 9-56 shows these results and lines of constant 10 to 1,000 j/cm^2 within which most pulsed lasers lie. The figure shows why heating and melting occur with long pulse or continuous wave lasers, and impulse delivery occurs with pulses shorter than approximately 10 μsec.

Damage thresholds for particle beams are more straightforward to calculate than those for lasers, because of the in-depth deposition of energy. We showed earlier that each particle in the beam loses energy at some rate, $\partial\epsilon/\partial x$, as it passes through the target. If the particle passes through a region of thickness, x, the energy lost within the region will be $(\partial\epsilon/\partial x)\Delta x$. Now the number of particles/cm^2/sec entering the region is simply

$$nv = I/\pi w^2 q \tag{92}$$

where n, v, I, w, and q are, respectively, the particle density, velocity, current, radius, and charge associated with the beam (see equation 30). Thus, the total rate of energy deposition wherever the beam passes through the target is

$$\dot{\epsilon} = nv\frac{\Delta\epsilon}{\Delta x} = \frac{I}{\pi w^2 q}\left(\frac{\partial\epsilon}{\partial x}\right)[\mathrm{w/m^3}] \quad . \tag{93}$$

A CPB deposits energy at a rate given by equation 93 within a long cylinder of radius w passing through the target. What are the orders of magnitude involved? In STP air a typical value of $\partial\epsilon/\partial x$, is approximately 4×10^{-16} j/cm if ionization is the dominant mechanism for energy loss. Since $\partial\epsilon/\partial x$ is proportional to target density, a comparable value in a solid target of density 10^3 times the density of air would be 4×10^{-13} j/cm. Therefore, a beam of protons of current I = 1,000 amphere and radius w = 1 centimeter would deposit energy at a rate

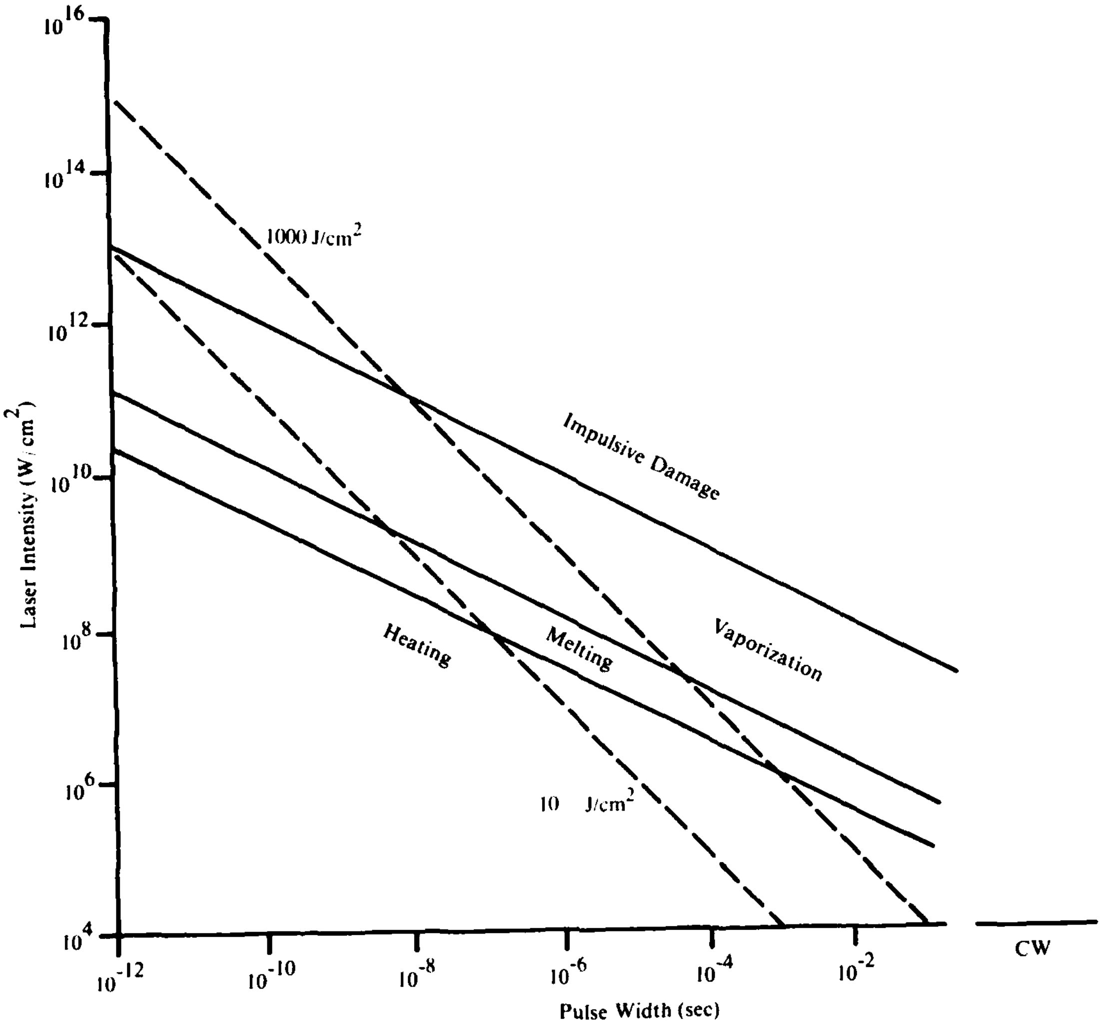


Figure 9-56. Target damage thresholds

$$\dot{\epsilon} = \frac{1}{\pi w^2 q}\left(\frac{\partial \epsilon}{\partial x}\right) = \frac{10^3\,\text{Amp}\,(4 \times 10^{-13}\,\text{j/cm})}{\pi(1\,\text{cm}^2)(1.6 \times 10^{-19}\,\text{coul})} = 8 \times 10^8\ \text{W/cm}^3 \quad . \tag{94}$$

At the rate given by equation 94, how long would it take to vaporize a hole through the target? If vaporization requires approximately 10^4 j/cm^3 (see equation 55), the time to vaporize the target would be approximately

$$t_v = \frac{10^4\,\text{j/cm}^3}{8 \times 10^8\,\text{W/cm}^3} = 1.25 \times 10^{-5}\ \text{sec}, \tag{95}$$

or approximately 10 μsec. This is a fairly short time to drill all the way through a target, in comparison with the 0.1 second required for a 10^4 W/cm^3 laser beam to drill its way through a target of 1 cm thickness.

Should we worry about thermal conduction with CPBs? After all, heat will flow radially out of the cylinder being heated by the beam, and will produce a temperature profile similar to the profile shown in figure 9-57.

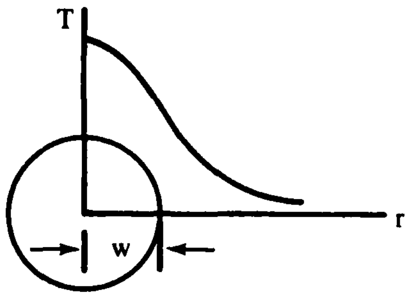


Figure 9-57. Temperature profile in a target heated by CPB.

The temperature in the center will rise until the flow of radial heat equals the rate of energy deposition from the beam. Energy is deposited within the beam volume and flows through its surface.

$$u = -K_c \operatorname{grad} T, \ (\mathrm{W/cm^2}) \tag{96}$$

where K_c is the thermal conductivity, gives the flow rate out of the surface (see equation 56). Thus a target of thickness, l, will reach a steady-state temperature when

$$\boxed{\underset{\frac{\text{Energy/Sec}}{\text{Volume}}}{\dot{\epsilon}} \quad \underset{\text{Volume}}{\pi w^2 \ell} \quad = \quad \underset{\frac{\text{Energy/Sec}}{\text{Area}}}{u} \quad \underset{\text{Area}}{2\pi w \ell}} \tag{97}$$

$$u = \frac{\dot{\epsilon} w}{2} \ . \tag{98}$$

A glance at figure 9-56 suggests that we estimate the temperature gradient in equation 96 as

$$\operatorname{grad} T \simeq -T/w \tag{99}$$

Substitution of equation 99 into equation 96 and 98 permits solution of equation 98 for T as

$$T = \frac{\dot{\epsilon} w^2}{K_c} \ . \tag{100}$$

By recalling that the thermal diffusivity, D, is K_c/C_p (see equation 59), we can express equation 100 as

$$T = \frac{\dot{\epsilon} w^2}{DC_p} \qquad (101)$$

What kind of T is predicted by equation 101? Using $\dot{\epsilon} = 8 \times 10^8$ W/cm³ from equation 94, and taking w = 1 cm, D = 1 cm²/sec, ρ = 1 gm/cm³ and C = 1 j/gmK, we get

$$T = \frac{8 \times 10^8 \text{ W/cm}^3 \text{ (1 cm}^2)}{(1 \text{ cm}^2/\text{sec})(1 \text{ j/gmK})(1 \text{ gm/cm}^3)} = 8 \times 10^8 \text{ K}$$

This is a temperature many orders of magnitude greater than the vaporization point of any substance. Therefore, the target will have vaporized long before thermal conduction can carry any energy away. Target damage is exceedingly simple with CPBs as opposed to lasers. We can use equation 93 to calculate the energy deposition rate, determine the energy necessary to achieve the desired level of damage, and ensure that the beam bears upon the target long enough to deliver the required energy. Without the propagation issues, this discussion of damage applies equally to neutral particle beams striking a target in space.

Summary

The following summary of key ideas can be helpful in estimating damage criteria for almost any proposed weapon system. However, we can use them to make specific estimates for lasers and particle beams as well. A good all-purpose damage criterion is to achieve an energy density of 10^4 j/cm³ within the target. Lasers deposit their energy on the target surface. A metallic surface may reflect a large fraction, approximately 90 percent of the incident energy. Thermal conduction carries the deposited energy into the interior of the target. The target will begin to melt, or vaporize, when the beam intensity exceeds $S_{m\,(or\,v)} = (D/t_p)^{1/2} \rho C\, T_m$ (or T_v). If S_v is exceeded, the target will be eroded at a velocity $V_d = S/L_v$. Above S_v, both mechanical and thermal damage is possible. CPBs deposit their energy throughout the target volume. A large fraction of the incident energy may pass through the target. The target will vaporize when the beam duration is such that $\dot{\epsilon} t_p > 10^4$ j/cm³. Thermal conduction is unimportant in CPB interactions.

BIBLIOGRAPHY

"Advanced Space Laser." *Spaceflight* (April 1970):161.
Allen, Leslie. *Essentials of Lasers.* Oxford, N.Y.: Pergamon Press, 1969.
"Applied Optics." *Proceedings of the IEEE* (October 1966).
Baker, Charles E. "Laser Display Technology." *IEEE Spectrum* (December 1968):39 50.
Benjamin, Robert F. and Gottfried I. Schappert. "Prepulse Damage to Target and Alignment Verification." *Journal of Applied Physics* 50 (January 79):7.
Birnbaum, George. *Optical Masers: Physical Principles.* New York: The Academic Press, 1964.
Bloom, Arnold L. *Gas Lasers.* New York: Wiley, 1968.
Brotherton, Manfred. *Masers and Lasers: How They Work, What They Do.* New York: McGraw-Hill Co., 1964.
Brown, Ronald. *Lasers, Tools of Modern Technology.* Garden City, N.Y.: Doubleday and Sons, 1968.
Burns, F. P. "High-Power Lasers: Their Performance Limitations, and Future." *IEEE Spectrum* (March 1967):115.
Callen, W. Russell, Donald C. O'Shea, and William T. Rhodes. *Introduction to Lasers and Their Applications.* Reading, Mass.: Addison-Wesley, 1977.
Carroll, John M. *The Story of the Laser.* New York: Dutton, 1970.
Carslaw, Horation S. and J. C. Jaeger. *Conduction of Heat in Solids.* London: Oxford University Press, 1948.
Davidson, Ronald C. *Theory of Nonneutral Plasmas.* New York: Benjamin-Cummings Press, 1974.
Eisberg, Robert M. *Fundamentals of Modern Physics.* New York: Wiley, 1967.
Elion, Herbert A. *Laser Systems and Applications.* Oxford, N.Y.: Pergamon Press, 1967.
Fishlock, David. *A Guide to the Laser.* London: MacDonald and Co., 1967.

"Giant Pancakes of Laser Light to Measure Precise Distance to Moon." *Space World* (Oct 1969):30-32.
Gray, E. E., ed. *American Institute Physics Handbook*. New York: McGraw-Hill, 1957.
Herriott, Donald R. "Applications of Laser Light." *Scientific American* (Sep 1968):140-154.
Jackson, J. D. *Classical Electrodynamics*. New York: Wiley, 1962.
Kittel, Charles. *Introduction to Solid State Physics*. 3d ed., New York: Wiley, 1966.
Lengyel, Bela A. *Lasers*. 2d ed., New York: Wiley, 1971.
Parmentola, John and Kosta Tsipis. "Particle Beam Weapons." *Scientific American* 240 (April 1979):54.
Robinson, Clarence A. "Air Force Emphasizes Laser Weapons." *Aviation Week* (30 October 1978):51.
Robinson, Clarence A. "Key Beam Weapons Tests Slated." *Aviation Week* (9 October 1978): 42.
Robinson, Clarence A. "Power Generation Key to Beam Weapon Development." *Aviation Week* (6 Nov 78):51.
Seigman, Anthony E. *An Introduction to Lasers and Masers*. New York: McGraw-Hill, 1971.
Robinson, Clarence A. "US Pushes Development of Beam Weapons." *Aviation Week* (2 Oct 78):14.
Wade, Nicholas. "Particle Beams as ABM Weapons: General and Physicists Differ." *Science*. 196 (22 Apr 1977): 407.
Wade, Nicholas. "Charged Debate Erupts over Russian Beam Weapon." *Science*.196 (27 May 1977):957-960.

Chapter 10

RELIABILITY OF SPACE SYSTEMS

The development of the digital computer has made possible a high degree of accuracy in computations for space research and engineering and has increased overall efficiency. In the operation of any space system, accuracy and efficiency are paramount. The same is true for any ballistic missile system. Since these systems are all very expensive, operational failure is extremely costly. In addition, the success of the system is always very important not only to the Air Force but also to the nation as a whole. Each space shot carries with it the prestige of the United States, and the very survival of the free world may rest upon the reliability of missile systems.

The advent of missiles and space systems has outmoded the "fly and fix" philosophy. Good systems must operate when fired. We must keep holds and mission aborts to a minimum. To accomplish this end, both manufacturers and operators of space systems must pay more attention than ever before to the reliability of the system as a whole. This requirement poses two critical problems, how to measure the reliability and how to increase the reliability. We will discuss both of these problems in this chapter. As a prelude, it is necessary to study the subject of probability, at least in its simplest form. Reliability is a term meaning the probability that equipment will perform a required function under specified conditions, without failure, for a specified period of time. Notice that reliability is a probability. It is a probability related to a system under specified conditions and for a specified time.

PROBABILITY

We can define the probability that a given event will occur as the number of ways in which the event can occur, divided by the number of ways the event can occur plus the number of ways the event can fail to occur.

If N_s is the number of ways in which a given event, A, can occur and N_f is the number of ways in which the event can fail to occur, then the probability of occurrence of the event is

$$P(A) = N_s/(N_s + N_f) \quad . \tag{1}$$

The probability of the event not occurring is

$$P(\text{not } A) = N_f/(N_s + N_f) \quad . \tag{2}$$

It is obvious from the definition of probability that values of probability range from 0 to 1. A 0 probability represents an impossibility, whereas a probability of 1 represents an absolute certainty.

We can find another important relationship from the definition of probability by adding equations 1 and 2. The equation resulting is

$$P(A) + P(\text{not } A) = 1 \quad . \tag{3}$$

However, we must meet two conditions before we can use the three equations given above. We must know the number of possible outcomes in a particular event, and we must know the possible outcomes that are favorable (or unfavorable) to the event.

We can illustrate probability by rolling a pair of unbiased dice. For example, what is the probability of obtaining a 7 in one roll? To answer this question, consider one roll as an event and then count the number of ways that the event can result in rolling a 7. The possible ways to succeed are to obtain the following combinations: 1 and 6, 2 and 5, 4 and 3, 3 and 4, 5 and 2, or 6 and 1. Thus, there are six ways to succeed. The total number of ways that the event can occur is 36, and the number of ways in which the event can occur and fail is 30. Therefore:

$$P_s = \frac{N_s}{N_s + N_f} = \frac{6}{6 + 30} = \frac{6}{36} = \frac{1}{6}$$

$$P_f = \frac{N_f}{N_s + N_f} = \frac{30}{6 + 30} = \frac{30}{36} = \frac{5}{6} \quad .$$

In situations dealing with a limited number of items, such as dice, cards, and coins, it is relatively simple to count the possible events and meet the above conditions. However, in many instances, the number of possible outcomes is too great to count, or the final result will depend upon the favorable or unfavorable outcomes of more than one event. In cases like these, we must use other equations and theorems to calculate probabilities.

Mutually Exclusive Events

We say that two or more events are mutually exclusive if one, and only one, event can occur at one time. In a mutually exclusive sequence, the probability that either one or another of the events will occur equals the sum of the probabilities of occurrence of the individual events. We can make a mathematical statement of this rule. It will read

$$P(A \text{ or } B) = P(A) + P(B) \tag{4}$$

We can extend this equation to include any number of mutually exclusive events. A simple problem can show the application of this rule very well. Consider a box containing one white ball, one red ball, and three black balls. What is the probability of drawing a red or white ball in one draw? We can effect the successful accomplishment of this endeavor in two ways: drawing a red ball or drawing a white ball. We can draw only one of these on any one trial. This is a case of mutually exclusive events. Consequently, the probability is the sum of the probabilities of drawing a red ball and drawing a white ball. From equation 1, these are one out of five in each case. Therefore, the probability of drawing either a red or a white ball is, by equation 4, $P(\text{red or white}) = 1/5 + 1/5 = 2/5$. We can find the probability of drawing a black ball as follows:

$$P(\text{red or white}) + P(\text{black}) = 1$$

$$P(\text{black}) = 1 - 2/5 = 3/5 \quad .$$

Contingent Probabilities

Suppose the undertaking is some large endeavor consisting of a number of separate events. If success or failure of the endeavor depends upon success or failure of each of the events in the endeavor, the probability of the endeavor depends on the probability of the events. If this is the case, we combine the probabilities for each of the events by multiplying them, not by adding them. We obtain the probability of success or failure for the endeavor by multiplying the probabilities of success or failure for each of the events in the endeavor. Suppose the endeavor is to launch a Saturn booster. Success or failure in this endeavor depends on the success or failure of a very large number of events, but we will simplify the

example by discussing the success of first-stage performance. If the probability of success in fueling is .9 and the probability of successful ignition is .8, what is the probability of success in the endeavor? Since success in the endeavor depends on success in both events, we can compute the answer by multiplying the probabilities of success for the two events. Thus, probability of a successful launch equals .9 times .8 or .72.

Sometimes, success in an endeavor may depend on failure in one or more of the component events. For example, to succeed in getting from the base to the bomb release line, an aircrew must fail to abort, fail to be shot down, and fail to make a gross navigational error. In some cases, success in the endeavor depends on success in some events and failure in other events.

We may classify events further as being either independent or dependent. Independent events are those in which the outcome of one event does not affect the outcome of any other event in the endeavor. Dependent events are those in which the outcome of one event does influence the outcome of other events in the endeavor.

Contingent independent (successful occurrence). The first example of contingent probabilities is contingent independent (successful occurrence). What is the probability of drawing two black balls from a box containing three black balls and two white balls, if we make two draws and we replace the ball drawn before the second draw? In this instance, the endeavor consists of two events, and the successful outcome of the endeavor depends on being successful in each event, that is, in drawing a black ball each time. The events are independent in that the first draw will not affect the second draw, since we will replace the ball. The probability of drawing two black balls is the product of the probabilities of drawing a black ball each time. The probability of drawing a black ball is three out of five in each instance. Therefore,

$$\begin{aligned} P(2\text{ black}) &= P(\text{black}) \times P(\text{black}) \\ &= 3/5 \times 3/5 \\ &= 9/25 \text{ or } 36 \text{ percent} \end{aligned}$$

Contingent dependent (successful occurrence). The second example is contingent dependent (successful occurrence). What is the probability of drawing two black balls from the box if we do not return to the box the first ball drawn? This problem is similar to the first example except that the events are dependent, since the probability on the second draw is dependent on the outcome of the first draw. The probability of getting a black ball on the second draw is either three out of four or one out of two, depending on whether or not we drew a black ball the first time. The probability of drawing a black ball the first time is three out of five. If the ball is black, then the probability that the second ball will be black is one out of two. The probability that both balls will be black is the product of these two probabilities:

$$\begin{aligned} P(\text{black then black}) &= 3/5 \times 1/2 \\ &= 3/10 \text{ or } 30 \text{ percent} \end{aligned}$$

Thus far, the examples have been such that the successful occurrence of an endeavor depended on the successful occurrence of the separate events. There are endeavors in which the nonoccurrence, or failure, of an endeavor depends on the nonoccurrence, or failure, of the separate events.

Contingent dependent (success and failure). This third example of probabilities illustrates contingent dependent (success and failure). What is the probability of drawing at least one black ball in two draws if we do not return the first ball drawn to the container? To draw at least one black ball, we do not need to draw a black ball on each attempt. Rather, not drawing at least one black ball in the endeavor depends on not drawing a black ball on each attempt. Since we did not replace the first ball, the events are dependent. We can solve the problem in the following manner:

$$\begin{aligned} P(\text{no black then no black}) &= P(\text{no black}) \times P(\text{no black}) \\ &= 2/5 \times 1/4 \\ &= 1/10 \text{ or } 10 \text{ percent} \quad . \end{aligned}$$

If the probability of not getting a black ball is one out of ten, then the probability of obtaining at least one black ball is $1 - 1/10 = 9/10$ or 90 percent.

Contingent independent (success and failure). The fourth example of probability illustrates contingent independent (success or failure). What is the probability of drawing at least one black ball from the container in two draws if we return the first ball drawn to the container? This example is similar to the examples of contingent dependent (success or failure) except that the events are now independent, since the first draw does not influence the second draw. For simplicity, we may define success as the act of obtaining a black ball. Then, failure is the inability to obtain a black ball and we denote this by P_f.

$$\begin{aligned} P(\text{no black, no black}) &= P_{f_1} \times P_{f_2} \\ P(\text{no black, no black}) &= 2/5 \times 2/5 \\ P(\text{no black, no black}) &= 4/25 \text{ or } 16 \text{ percent} \quad . \\ P(\text{at least one black}) &= 1 - .16 \\ &= .84 \text{ or } 84 \text{ percent} \quad . \end{aligned}$$

We can combine the preceding as follows:

$$\begin{aligned} P(\text{at least one black}) &= 1 - P(\text{no black, no black}) \\ &= 1 - (P_{f_1} \times P_{f_2}) \quad . \end{aligned}$$

Since P_{f_1} equals P_{f_2}, we can write this as

$$P(\text{at least one black}) = 1 - (P_f)^2 \quad .$$

For the general case, we can write the equation as

$$P(\text{at least one A}) = 1 - P(\text{no A})^n \quad . \tag{5}$$

Equation 5 gives the probability of at least one occurrence in a series of n repeated trials. We can use equation 5 to calculate the number of trials necessary to achieve a given probability of at least one occurrence. In space systems, as we will explain later, we can use it to calculate the reliability of redundant systems.

Minimum probability. The fifth example illustrates minimum probability. How many draws must we make from a container of three black balls and two white balls to have a minimum probability of 90 percent of obtaining at least one black ball? We can solve the problem by using equation 5.

$$\begin{aligned} P(\text{at least one black}) &= 1 - P(\text{no black})^n \\ .90 &= 1 - (.4)^n \\ -1 + .90 &= -(.4)^n \\ .10 &= (.4)^n \\ n &= 2.51 \quad . \end{aligned}$$

Therefore, it would take three draws to have a minimum probability of 90 percent of drawing at least one black ball.

Guide for the Solution of Probability Problems

We have included the following step-by-step guide for the solution of probability problems to make it easier for you, the user, to organize your problem solving. If used as recommended, without skipping steps, you should find it much easier to solve probability problems.

Step 1. Define success for the problem under consideration.

Step 2. Does success in the endeavor depend upon the outcome of a series of events?

- If no, compute the single probability and you have solved the problem.
- If yes, go to step 3.

Step 3. Are these events mutually exclusive? Does being successful in one way rule out being successful in any other way?

- If yes, determine the probabilities of success for the several mutually exclusive events that may happen. The sum of the probabilities of success of the several events equals the probability of success in the endeavor, and you have solved the problem.
- If no, go to step 4.

Step 4. Does success in the endeavor require success in each event?

- If yes, determine the probability of success in each event. The product of the probabilities of success in each event equals the probability of success in the endeavor and you have solved the problem.
- If no, go to step 5.

Step 5. Does success in the endeavor require nonsuccess in each event?

- If yes, determine the probability of failure in each event. The product of the probabilities of failure in each event equals the probability of success in the endeavor, and you have solved the problem.
- If no, go to step 6.

Step 6. Does success in the endeavor require success in some of the events and failure in others?

- If yes, determine the probabilities of success for those events where success in the endeavor depends upon success in the events; determine the probabilities of failure for those events where success in the endeavor depends upon failure in the events. The product of these probabilities is the probability of success in the endeavor, and you have solved the problem.
- If no, then failure in the endeavor must depend on failure in each of the events. Then, it is necessary to determine the probability of failure in each event. The product of the probabilities of failure in the events equals the probability in the endeavor. You can find the corresponding probability of success by using the following formula: $P_s = 1 - P_f$ (where P_s = probability of success in the endeavor and P_f = probability of failure).

RELIABILITY

Suppose mission support personnel have fueled and checked a booster for a spacecraft, and

the countdown is complete. All systems are "go." Will the booster and its spacecraft actually perform their mission without failure for the specified period of time? Actual experience shows that, under these conditions, some do, and some do not. However, before the actual launch, the directors of a project want a good estimate of the chances of success, or of the reliability of the system.

Suppose a booster is in the process of design, development, and production. With the thousands of component parts that must be assembled into an operating system, the assemblers will make some mistakes both in the design and in the process of production to meet design specifications. What can they do to measure the reliability of a design and of the component parts that enter into the finished product? Reliability, which is a word that is becoming more and more associated with both space systems and missile systems, is a probability idea. As we stated earlier, it is the probability that a system will perform a required function under specified conditions, without failure, for a specified period of time. We can apply this idea to a complex system consisting of a multistage booster, a spacecraft with more than one stage, and a recovery system. We can apply it to one stage of a system or one component part, such as a transistor or a valve.

Calculating the Reliability of a System from the Reliability of its Parts

Even though a system may have as many as 30,000 parts, assume a simple system consisting of 10 parts for purposes of illustration. Assume that we have measured the reliability of each part and found it to be .90. If the system is to operate, each of the parts must operate. Therefore, the probabilities of success are contingent upon the successful operation of each part. What is the reliability of this simple system? Clearly, it is $(.90)^{10} = .35$.

At first glance, .90 seems like a rather good reliability. However, if the reliability is no greater than .90, even a 10-component system becomes much too unreliable. A 30,000-component system would have ridiculously low reliability. What can be done? First, the component parts of a system must have reliabilities much higher than .90. Assume a 10-component system in which each part has a reliability of .99. Now the reliability of the simple system is $(.99)^{10} = .905$. Suppose the reliability of the parts is .999. Now the reliability of the 10-part system becomes $(.999)^{10} = .991$.

Now think of a system with 10,000 parts. One important conclusion from the simple example is immediately apparent. If reliability in a real system is to be around .90, the reliability of the parts must be very high indeed. Scientists and engineers must design and test, redesign and retest, all of the parts until their reliability approaches 100 percent. In a recent symposium on reliability, engineers talked about the reliability of parts of .9999 and higher.

Methods for measuring. Sometimes, engineers can compute the reliability of the equipment by first measuring the failure rate. For example, they might test a part for a booster by selecting a number of parts at random and starting them to operate under conditions that approximated the actual conditions of expected operation as closely as possible. Suppose that one part failed each 100 hours of operation on the average. Then, the failure rate, f, would be 0.01 per hour.

Another, and more common, procedure is to measure the mean time before failure (MTBF) and calculate the failure rate from this. For example, suppose the first part failed to operate after 40 hours, then more and more of the parts began to fail, but the last sample did not fail until 300 hours had passed. From this, we can calculate easily the mean, or average, time of operation before failure. Suppose it is 100 hours. Then $f = \frac{1}{\text{MTBF}} = \frac{1}{100 \text{ hours}} = .01$ per hour.

If we know the failure rate and the desired operating time of the part, we can calculate the probability that each part will operate for this time by using the formula $P_s = e^{-ft}$ (where P_s is reliability, f is failure rate, and t is time of desired operation). For example, if a piece of equipment had a failure rate of 0.001 per hour and we expected it to operate for ten hours, its reliability would be $e^{-(0.001)(10)} = e^{-.01} \approx 0.99$.

Advantages and disadvantages of the procedure. Industrial firms widely use the procedure

previously outlined and numerous modifications of it. It has many advantages. In the first place, it enables the engineers to identify parts that have a reliability that is too low. Then, they can redesign the part, test it again, and improve the product through repetition of the procedure. Further, it is not too complicated a procedure. It involves only careful testing and some calculation. However, are the results truly valid for testing the reliability of a complex system? Can we do this by testing individual components and calculating contingent probabilities?

The question focuses attention on some of the disadvantages of the system. The procedure assumes that the reliability of each component used in the calculation is constant. This may or may not be true. This is not true if the part has some minor modifications made after proper personnel have determined the reliability. The modification may either increase or decrease the reliability. This is not true after the part has been operating so long that it is beginning to wear out. If this is the case, the reliability will be falling, perhaps quite rapidly. Thus, the procedure is valid only if the parts are operating in that portion of their lives in which the failure rate is constant. This procedure is valid only if the test environment exactly duplicates the operational environment. Often, the exact operating conditions are very difficult to simulate because frequently one subsystem in a booster affects another subsystem near it in a manner that is difficult to predict. Finally, the procedure relies upon a process of sampling. The parts tested to the point of failure are not the actual parts that will be used in the booster or space system. However, we assume that, because the parts tested are a valid sample from those the spacecraft actually will use, inferences from the sample to the population from which they were drawn, are valid. This may or may not be true. However, in spite of its disadvantages, experts widely use the procedure of testing the reliability of a system by testing the reliability of the parts. Generally, they concede it will yield conservative figures. If we must make errors in estimation, it is better to err on the conservative side.

Improving a Low Reliability

If the reliability of either a system or a component is too low to be acceptable, there are many things that can be done. In fact, most large manufacturing companies have a staff of engineers and statisticians whose sole job is to study the reliability of the company's products and recommend procedures to improve reliability. This requirement within a company is becoming more important because the government has stated that system reliability must be one of the design *specifications in* its contracts.

Preventing infant mortality failures. The parts produced by a given piece-part production line normally are put into service over an extended period of time. However, if all of the like parts (transmissions, diodes, and so forth) were put into service simultaneously, and their individual failures plotted against time, the resultant histogram would be similar to figure 10-1.

The relative numbers are not important; the general shape of the histogram is. It tells us that a large number of failures occur immediately after we put the parts into service (infant mortality), and then the failures drop-off until the remaining parts begin wearing out.

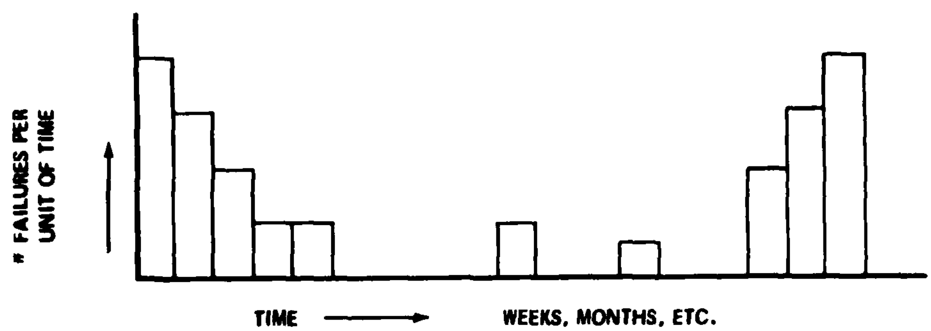


Figure 10-1. Piece-part failures.

To cull out these early failures, engineers must perform extensive screening tests in very severe conditions. Common aerospace industry tests include testing throughout the piece-parts range of operation, testing at very high and low levels of vibration, and complete X-ray examinations of each part.

Usually, this testing increases the cost of the parts as much as 10 to 20 times the original cost, depending on the degree of testing required. We must take care to avoid overtesting lest we encounter premature end-of-life failures.

Providing reliability through quality control. In the process of manufacturing a part, a shop foreman may look at a design and decide that if he made a small change, it would be easier to fabricate and would be as good. When the foreman adopts the change, often the part turns out to be just as good only in the sense that it will work. However, the change may have degraded seriously its reliability. To overcome this natural tendency of workers and foremen to try to improve a product, when they are not in a position to know whether a change is an improvement or not, many companies forbid any changes at levels of management beneath that of the design engineers. Configuration management is the name for this process of controlling changes.

Another type of quality control action consists of correcting the faulty design of a part that tests at a low reliability. This should be the work of the design engineer, not of the shop worker. In making the correction, the company must consider the part not only as an entity but also as an integral part of a larger operation.

Sometimes a part functions with high reliability only if it is absolutely clean. Dust, dirt, grease, hair, or other debris, even in the smallest quantity, may cause the part to fail. Examples of such parts are valves, pumps, and lines used to handle liquid oxygen. When foreign particles seriously affect reliability, often a company must use elaborate precautions in the manufacturing and handling of the part. The company may hold the workroom under positive atmospheric pressure so that no outside air will seep in. The company ensures the filtering and conditioning of all incoming air. At the same time, all workers wear gloves and special clothing. They must handle the part with special tools. The company has devised special cleaning procedures that are enforced rigidly. Finally, the company controls and inspects packaging very carefully. In fact, the workers undertake the whole process in an environment as clean and sterile as that of the hospital operating room.

Often there is a relationship between poor reliability of a part of a subsystem and lack of employee discipline. Humans become careless. Quite naturally, a few mistakes are made. In the manufacture of space and missile systems mistakes and carelessness are too costly. Most companies not only have training programs that impress upon workers the need for good working discipline, but they have continuing programs that encourage attention to detail and a sense of responsibility as well.

Quality control is in itself a large subject. We commented on it because it is one of the most important methods for improving the reliability of a part, subsystem, or system.

Providing reliability through testing. While the company intends for piece-part screening to ensure that they install only good parts into a subsystem, and for quality control procedures to ensure that the parts are made and installed correctly, the real test of success is whether the parts work as a unit in the overall space vehicle. The aerospace industry, recognizing that they cannot fix spacecraft on-orbit, has instituted an extensive series of unit, subsystem, and total system tests.

Whether a black box is a guidance computer, telemetry set, or other part, the company subjects each box to testing through or beyond its full range of operation and at the extremes of temperature, pressure, vibration, and acoustics that it will experience during actual flight. Only after successful completion of these tests will the company install the unit in the next higher level system.

Further, the first unit or subsystem off a new production line is normally a qualification test unit, or qual unit. The company will overstress the unit in all of the above areas to ensure that they overdesign it. After successfully completing the qual test, the company will declare that the production line units are flight qualified.

The complete systems test is the final testing done on a spacecraft. It is in keeping with the Air Force's factory-to-pad concept of operations. During systems test, the company will run the entire

spacecraft through all of its operations while still at the factory. In some cases the company will install the spacecraft in a large thermal-vacuum chamber so they can simulate the temperature and pressure extremes of space.

Testing of one-shot devices. A special problem arises when we need to know the reliability of one-shot devices. One-shot devices are items that we can only use once and we must test destructively. An ordinary match is a simple example. Ordnance devices such as squibs, igniters, or explosive bolts are common examples in space systems.

The procedure commonly used involves testing a certain number of random samples from a batch, and drawing inferences as to the reliability of the lot. The manufacturer fires home the units under conditions that closely approximate operating conditions, and then makes inferences to what can be expected when the user fires the remainder. That is, the manufacturer samples the inventory, tests the sample, and then makes inferences to the whole population from which the manufacturer drew the samples. Clearly, if the references are to be valid, the sample must be truly representative of the population. One way of doing this is to select a random sample and arrange the inventory such that every possible sample has an equal chance for selection. Suppose an attempt is made to fire a sample of 20 units. In the firing, 10 are successes and 10 are failures. The reliability of the sample is 50 percent. Now, inferences must be made from this sample to the whole population. We cannot simply say that the reliability of the population is 50 percent because chance might have decreed that the sample had a high proportion of either good or bad units in it. In another example, perhaps only nine would fire. In this event, the reliabilty of the sample was only 45 percent. Perhaps 11 of them would succeed with the reliability of the sample being 55 percent. How can an inference be made from the sample to the whole population?

Statisticians have worked out the way to make inferences, and table 10-1 summarizes their results. To check on results from firing 20, pick out a sample of 20 in the table. Then select from the lefthand column the number 10 as the number observed to fire successfully. At the intersection of the column and the row is the number .31. Since the statisticians built the table for a 95 percent confidence level for a reliability, this number means that if 10 units in the sample of 20 succeed, there would be 95 percent confidence in the reliability of the population being greater than 31 percent. Another way to express this is to say that, if 10 succeed out of a sample of 20, the odds are 19 to 1 that the true reliabilty is greater than 31 percent.

Table 10-1.
95% Confidence Levels for Reliability, p*

Number of Successes, S	1	2	3	4	5	6	7	8	9	10	15	20
0	0	0	0	0	0	0	0	0	0	0	0	0
1	.05	.02	.02	.01	.01	.01	.01	.01	.01	.01	0	0
2		.22	.14	.10	.08	.06	.05	.05	.04	.04	.02	.02
3			.37	.25	.19	.15	.13	.11	.10	.09	.06	.04
4				.47	.34	.27	.22	.19	.17	.15	.10	.07
5					.55	.42	.35	.30	.25	.22	.14	.10
6						.61	.49	.40	.35	.30	.19	.14
7							.65	.54	.45	.39	.24	.18
8								.69	.58	.49	.30	.22
9									.72	.61	.36	.27
10										.74	.42	.31
11											.49	.35
12											.56	.39
13											.64	.45
14											.72	.49
15											.82	.54
16												.60
17												.66
18												.72
19												.79
20												.86

*Based on Binomial Distribution

Other tables are available for other confidence levels, even as high as 99 percent confidence. However, we will continue to use the 95 percent table for illustration. Another way of looking at the situation is that if we repeated the sampling process many times, each time taking a sample of 20 units and each time using the table to make inferences from the sample to the population, the result would be right 95 percent of the time. There would still be a 5 percent chance that the inferences would be wrong.

Use of redundancy. Another method of improving reliability is the use of redundancy. Designers may design certain critical parts of a system so that two, or more, alternate or redundant parts are provided. They combine these parts in such a way that the system fails only if both, all three, or all four of the parts fail. In other words, the designers combined the parts in such a manner that the system will function if at least one of the redundant parts operate properly.

The calculation of reliabiilty for a redundant portion of a system is exactly the same as the calculation we used in the illustration of contingent independent (success and failure) under contingent probabilities. Suppose a company has expended all the effort under product control, and still a part has a reliability of only .90. The engineers may decide that they will use two such parts, combined in parallel, to improve the reliability. What is the probability that at least one of these parts will function as planned? Clearly, success of the system does not depend upon success of both parts. The probability of failure for each part is $1 - .90 = .10$. The probability that both parts will fail is $(.10)^2 = .01$. The probability that at least one part will operate is $1 - .01 = .99$. Thus by use of a two-path redundant subsystem, they have improved the reliability of the part from .90 to .99.

There are disadvantages to this plan. Using two parts instead of one increases weight, an item of great importance in space and missile systems, and increases cost. However, it is a plan often used because it has many advantages already mentioned. If a two-path redundant system is good, why not use a three-, four-, or five-path redundant system? Usually, this is not done because the weight and cost penalties increase with redundancy. Further, the gain follows a law of diminishing returns. Using two parts in parallel instead of one part increased the reliability by 9 percent in the example we gave. However, using three parts in parallel instead of two increased the reliability by only 0.9 percent. There may be circumstances under which a gain of 0.9 percent will compensate for the weight and penalties incurred. Thus, the decision to use redundancy and how much redundancy to use is a matter of judgment on the part of the design engineer.

Results. On programs that require many piece-parts, we must make a trade-off among several factors. We must consider the high costs associated with extensive screening, testing redundant systems, or acceptance of a certain risk by reducing the amount and/or degree of severity of the testing to reach a cost level compatible with the available funding. This can be a very critical decision, and the total success or failure of a multimillion dollar program may depend on where we make the compromise.

Figure 10-2 is a graph of the success (in percentage) of our US launch vehicles since 1957. While the data are only for launch vehicles and do not include the spacecraft, the success of the reliability programs instituted by the government and industry is clear. In fact, the major spin-off of the US space program may be the reliabilty techniques that it developed.

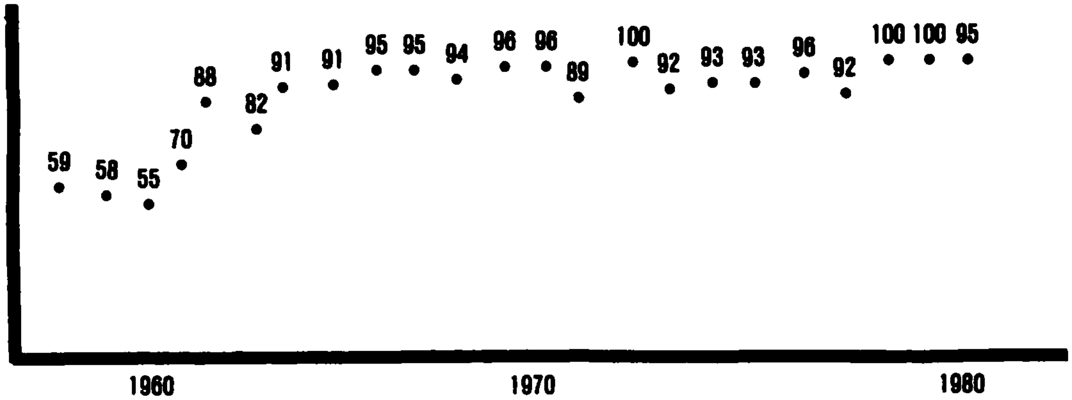


Figure 10-2. US launch success rate.

BIBLIOGRAPHY

Fraser, James A. "Reliability." *Air University Review*. 16, no. 6 (September-October 1965):23 28.
Middendorf, William H. *Engineering Design*. Boston: Allyn and Bacon, Inc., 1969.
Taylor, George A. *Managerial and Engineering Economy*. New York: D. Van Nostrand Company, 1964.

Chapter 11

BIOASTRONAUTICS

The broad scientific field of bioastronautics had its origin many years ago in the daring adventures of men who left the surface of the earth to explore the atmosphere surrounding the earth. Early in these attempts, people learned that flight in the atmosphere presented hazards to life never before encountered. Balloons descended with dead crew members and experimental animals. Later, pilots of airplanes were found dead in the wreckage of their crafts. There were many speculations as to the causes of these tragedies, but little was known other than the fact that high altitude presented a lower barometric pressure and a corresponding reduction of oxygen and nitrogen partial pressures. If travel in the atmosphere were to progress, we needed to learn a great deal about the cause and effect of the atmospheric environment. At this time, the medical services and physiological and biological researchers launched extensive investigations to determine the characteristics of the environment and the effects of this environment on people. In their research, they relied in part on the work of Claude Bernard, a famous physiologist, who, in 1878, wrote:

> The higher animals really have two environments an outside environment in which the organism is situated and the inside environment in which the tissue elements live. The living organism does not really exist in the outside world (i.e., in the atmosphere if it breathes) but in the liquid inside world formed by the circulating organic liquid which surrounds and bathes all the tissue elements. The internal environment surrounding the organs, the tissues and their elements, never varies; atmosphere changes cannot penetrate beyond it, and it is therefore true to say that the physical conditions of environment are unchanging in a higher animal. Each one is surrounded by this invariable world, which is, as it were, an atmosphere proper to itself in an ever changing world outside. Here we have an organism which has enclosed itself in a kind of hothouse. The perpetual changes of external conditions cannot reach it. It is not subject to them but is free and independent. All the vital mechanisms, however varied they may be, have only one object, that of preserving constant the conditions of life in the internal environment.

Claude Bernard was saying that the cell (the unit of living organisms) as found in tissues and organs (muscles, heart, lungs, and liver, among others) of higher animals must remain in a relatively stable environment, regardless of the external environmental changes surrounding the organism. We know that the external environment lies outside people and includes such influences as air pressure, humidity, temperature, odors, vibrations, noise, and acceleration.The internal environment consists of conditions within the body that affect the activities of cells, tissues, and organs. These internal conditions include such influences as body temperature, blood oxygen, carbon dioxide in the blood, and blood chemistry in general. Human beings can withstand moderate changes in the external environment without any deterioration in performance primarily because their internal environments are kept reasonably constant by regulatory mechanisms controlled by the central nervous system. These mechanisms provide for exchanges with the external environment to maintain body temperature and blood chemistry, among other things, within rather narrow limits. This whole process of maintaining a relatively constant internal environment is called homeostasis.

In certain instances, changes in the external environment can prove too extreme for the regulatory mechanisms. In such cases, these mechanisms are unable to maintain a constant internal environment, and there results a deterioration in the performance of the sense organs, central nervous system, and/or muscles and glands. When the human being is a part of the man-machine system, the

performance of the system deteriorates as well. This is essentially what the early aviation physiologists and aviation medical personnel had to work within the infancy of their endeavor. Since their beginning they have amassed a large volume of information designing or outlining all of the known parameters of the external environment and their effects on a person's internal environment and regulatory mechanisms. They have detailed definite tolerance limits to external forces for physiological functions in the human.

Therefore, we have known values or quantities of physiological functions that we can apply to a known value or quantity of external environmental stress. If the value of the physical stressor exceeds the ability of the body to adjust to it, we must reduce the stress or provide some support to the regulatory mechanisms of people for them to tolerate the stress and remain homeostatic.

Now that we have entered the new space age and people are attempting to fly in the hostile environment of space, again we look to the psychological research field for help in meeting the stresses placed on people by travel in this new environment. This field of study is called bioastronautics, which is the study of the physiological and psychological problems facing people in space travel.

Bioastronautics is a generic term that encompasses several related disciplines. One convenient breakout divides bioastronautics into the three fields of bioresearch, bioengineering, and bioenvironmental efforts. Bioresearch includes biomedicine and biology. Biomedicine is the study of the adaptiveness and tolerance of people to space operations. Biology is the study of living organisms, for example, insects, plants, fungi. Bioengineering is divided into bioinstrumentation and people/system integration. The aim of the former is to develop the appropriate instruments for life sciences research. The aim of the latter is to determine the optimum uses of a person's capabilities in space missions. Finally, bioenvironmentalists develop the technology for highly reliable life support and protective systems to sustain people on space flights. Bioastronautics is not the exclusive domain of any one field or skill. It covers a wide category of specialties and requires the disciplined, integrated efforts of engineers, physicians, research physiologists, technicians, and support personnel to succeed.

The problems are many, but, thanks to the diligent research of physiologists and medical personnel, some of the answers to these problems in space flight were available long before humankind's first orbital flight. These answers were based on the data gathered in the preceding 50 years of aviation medicine and physiological research. Thus, although the name bioastronautics is new, the techniques and procedures used to determine people's tolerance to external stress are well known, and space flight by people is relatively predictive and safe as a result. To understand these parameters more fully, we must look at the major physiological stresses placed on people during space flight.

PHYSIOLOGICAL STRESSES

As mentioned earlier, changes in the external environment may at times be very severe and too extreme for a person's regulatory mechanisms. These changes cause stress on these mechanisms, which is nothing more than pushing the homeostatic levels of people to the tolerance limits or beyond. As people venture into space, they are exposed to a series of stresses, each of which could prohibit manned space flight if no corrective measures are available. This section discusses most of the major environmental stress parameters, how they affect people, and how such forces are overcome so that people can maintain bodily functions within homeostatic limits. In general, there are two types of environment in which changes can cause severe stress — the physical and the mechanical. Changes in atmospheric conditions cause stresses in the physical environment and operation of the space vehicle causes stresses in the mechanical environment. We will discuss each of these environments separately.

Physical Environment

An atmosphere approximately 100 miles thick surrounds the earth. This atmosphere provides ambient pressures, temperatures, humidity, and certain gases required in our normal environment. Also, it provides a shield which prevents certain matter and radiant energy from reaching the earth's surface.

As people progress from the surface of the earth upward through the sea of air, the atmospheric

pressure, gas concentrations, and humidity decrease and the temperature varies widely. Above the atmosphere, people are no longer protected from meteorites and various forms of radiant energy that we find in space. These changes in the external physical environment are too extreme for our regulatory mechanisms and we cannot exist without a variety of protective devices.

At extreme altitudes and in space, people require a sealed cabin or suit that provides an adequate partial pressure of oxygen (O_2) for tissue oxygenation. The cabin may contain nitrogen, which makes up 80 percent of the earth's atmosphere. Early US space vehicles contained 100 percent O_2 at 5 pounds per square inch, simplifying gas control and capsule engineering. Absence of nitrogen eliminated the possibility of altitude sickness, where dissolved inert gas in tissues forms bubbles bringing on severe medical consequences. The disadvantages of the five pounds per square inch O_2 atmosphere include increased fire hazards and a tendency to produce anemia. There is evidence that exposure over extended durations of 5.0 pounds per square inch (284 millimeters of mercury Hg) pure oxygen may prove to be toxic. To avoid the toxic effect of 100 percent oxygen at 5 pounds per square inch, Skylab used a 70 percent oxygen and 30 percent nitrogen mixture to provide an oxygen partial pressure equivalent to 20 percent at one atmosphere. The space shuttle is designed to operate at 14.7 pounds per square inch (one atmosphere) with 80 percent nitrogen and 20 percent oxygen.

Heat and humidity. A person's internal heat regulatory mechanisms tend to keep the body at a constant temeprature of 98.6° F. Normally, the body loses heat constantly through the lungs and skin. However, if a person needs to lose more heat to maintain constant body temperature, the regulatory mechanisms come into play to increase the heat loss. These mechanisms include dilation of the blood vessels near the skin (bringing more blood near the surface of the skin) and perspiration, which increases heat loss due to evaporation. Conversely, if a person's environment cools and the person needs to conserve heat, perspiration stops and the blood vessels constrict. In addition, if the cold becomes extreme, reflex shivering causes the muscles to produce more heat. Normally, a nude body can maintain thermal balance easily if the environmental temperature is 70° to 80° F and the relative humidity is approximately 45 percent.

In spacecraft operation, many factors come into play. The flux of heat energy is large, and the regulatory mechanisms employed by people to maintain body temperature are inhibited. The sources of heat energy, other than from people themselves; include the electronic equipment in the vehicle, friction heat as the vehicle leaves or reenters the atmosphere, and heat energy from the sun, which an atmosphere no longer dissipates before it reaches the vehicle. Because the heat sources produce a high heat load and a person's normal heat loss mechanisms are reduced, people require additional protection. Special suits are worn under the pressure suit to provide circulating water to help dissipate the heat, and air conditioners provide additional protection by regulating the ambient cabin temperature.

Contaminants. Contaminants, such as carbon dioxide and methane, among others, become a problem in space capsules due to the confined space and the sealed characteristics of the vehicles. We must consider all atmospheric substances to be toxic if introduced into the body in amounts greater than some threshold value. Therefore, we must treat each contaminant individually in terms of its own threshold value. In all cases, both concentration and the time of exposure are critical conditions. The scope of this chapter does not cover the multitude of contaminants that are possible in a space capsule. Therefore, we may say that we must keep all substances normal to a space cabin environment at nontoxic levels even to the point of removing some substances entirely.

Radiation. We discussed the types of radiation that people encounter in space in the chapter on space environment (chapter 1). This chapter discusses the hazards to people from the wide range of radiant energy to which they will be exposed when they leave the protective blanket of the earth's atmosphere. The ultraviolet, visible, and infrared radiations found in space could cause problems for the members of the crew if they were unprotected. However, the spacecraft, their suits, and protective visors have eliminated almost entirely all hazards from these radiations.

The main hazards to life are the ionizing radiations and their effect on living cells. A consideration

of the physical changes that occur within the cell can explain the damage to a living cell. We know that radiant energy can alter the normal structure and the electrochemical makeup of biological material. This results in the breakdown of the molecular structure and/or the production of toxic substances. Usually, these changes are lethal to cells, especially if the enzymes, which are catalysts for all biological reactions, are destroyed. If the cell chemistry is disrupted, the cell dies. If sufficient cells of an organ are damaged, the organ will cease to function and the person will die. A summary of the biological effects of radiation follows. The effects may be direct (direct damage to the cell nucleus) or indirect (damage to the cell's enzymes and chemistry); immediate or delayed depending on the kind of damage; reversible and irreversible depending on the amount of radiation and type of tissue involved. The three main organ systems sensitive to radiation are the blood system, digestive system, and central nervous system.

One unit of measurement used in relating radiation effects in biological materials is the radiation absorbed dose (RAD). One RAD is equivalent to the absorption of 100 ergs (units of work) of energy per gram of living tissue. Figure 11-1 shows the known effects of acute whole-body radiation in RADs. If a population receives approximately 200 RADs of acute, whole-body radiation, up to 60 percent of those individuals will become ill within three hours. However, the tolerance differences between individuals is quite wide, and the rate of onset of illness can vary considerably. If the dose received is approximately 450 RADs, up to 50 percent of the population will die and the other half will become ill. What kinds of doses can we expect in normal space operations? Figure 11-2 shows the expected dose levels to people in a space vehicle with 8 to 10 grams/cm^2 shielding. Two factors are apparent. The dose levels are rather small and the rate of onset of the dose levels are spread out over days, weeks, and months.

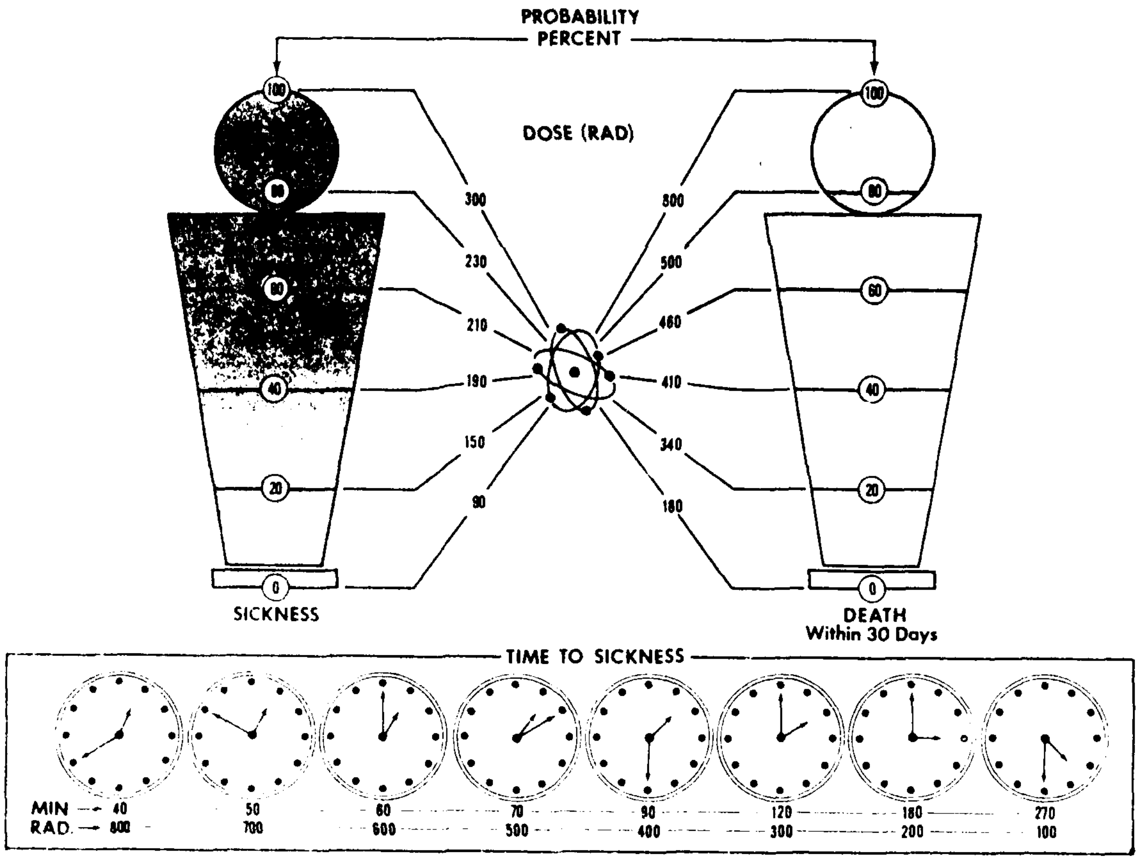
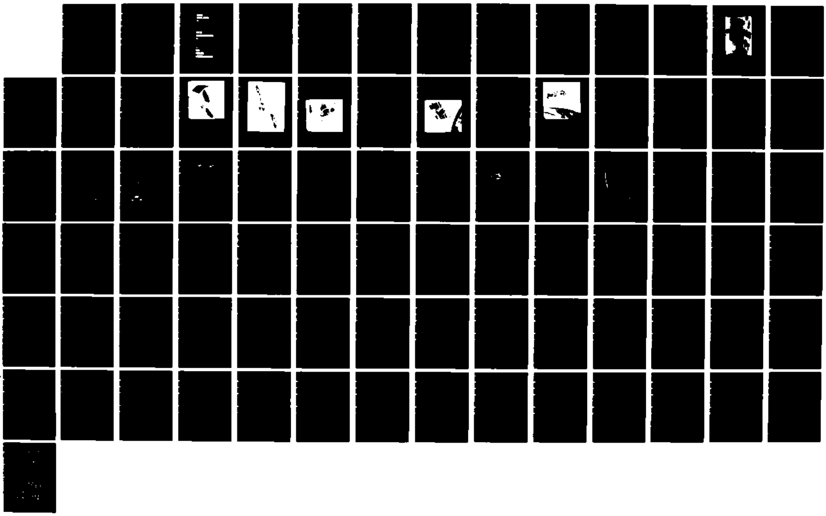
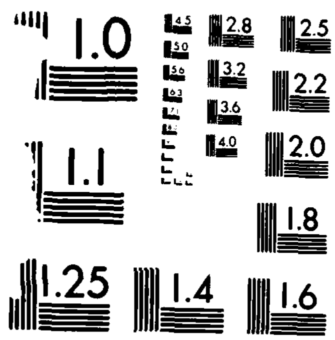


Figure 11-1.

AD-A165 038 SPACE HANDBOOK(U) AIR MINISTRY LONDON (ENGLAND) 4/4
C D COCHRAN ET AL. JAN 85 AU-18

UNCLASSIFIED F/G 15/7 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DOSE ACCUMULATION - RAD

ZONE	8 - 10 gm/cm² SHIELDING	TRAINING 1 - 3 DAYS	EXPLORATION 1 - 2 WKS	2 - 4 WKS	APPLICATION 1 - 3 MONTHS
1	LOW EARTH ORBIT	0.06 RAD	0.28 RAD	0.56 RAD	1.8 RAD
2	ANOMALY	0.5 "	7.0 "	14.0 "	45.0 "
3	2 PENETRATIONS OF NATURAL BELTS	10.0 "	10.0 "	10.0 "	10.0 "
5	GALACTIC BACKGROUND	0.045"	0.2 "	1.0 "	3.0 "
	TOTAL	~11.0 "	~18.0 "	~26.0 "	~60.0 "
4	SOLAR FLARE	100 RAD	100 RAD	100 RAD	100 RAD

Figure 11-2.

Based on this information, Col J. E. Pickering, US Air Force, Medical Corps, has proposed mission selection according to exposure rates. His proposal considers the age of the crew member, the number of years which the candidate expects to spend in the flight program, and the provision of adequate recovery periods after each exposure. He assumed that space programs would continue to use a space-qualified crew member for as long as the program could assure the safety of the crew member. In this approach, the program would permit an older person (35 years and older) only one year of flight. The schedule would call for one week of recovery for each RAD received by the crew member. In the case of an older crew member, the radiation exposure would need to be very carefully managed to gain maximum usage of the crew member's flight time through training, exploration (space missions), and applied (routine repeated flights) phases. A high rate of exposure during training or exploration would deny the crew member's experience for routine applied missions. In the case of younger people who will have three to five years of flight available, proper management of their radiation exposure will provide a long, useful career in the exploration and applied phases of operation. Colonel Pickering proposed that a younger crew member could tolerate a total dose of 80 RADs distributed over a five-year career as long as the program provided a compensatory one-week recovery period time for each RAD received after each period of exposure. This suggested dosage still permits an additional emergency dose of 100 RADs due to a solar flare without serious jeopardy to mission success. This seemingly large total dosage is considered reasonable because the numbers of the population exposed over the next several years will be small and carefully observed. Thus, the added risk due to radiation above and beyond the total risks associated with the mission will remain extremely small. Colonel Pickering's reasoning behind the approach is to return flexibility to the operational mission as far as radiation is concerned (see fig. 11-3, 11-4, and 11-5).

Mechanical Environment

The problems associated with the mechanical environment include all those stresses placed on an astronaut as a result of the vehicle and its operation in the space environment.

Acceleration. As the rocket engines thrust the space vehicle toward orbital velocity and as the vehicle changes velocity during reentry, it experiences a significant increase in acceleration forces. We measure these forces in units of gravity forces (g) with one g being equivalent to the force of gravity acting on a body at sea level. Booster accelerations are unique for each stage and for each type of vehicle. The peaks of acceleration enroute to orbit range from three to eight g's. Figure 11-6 shows the acceleration forces for a typical manned vehicle flight profile.

Due to the effects of increased forces on the cardiovascular and the muscular systems, an astronaut must have some protection. The circulatory system tolerance varies, depending on the manner in

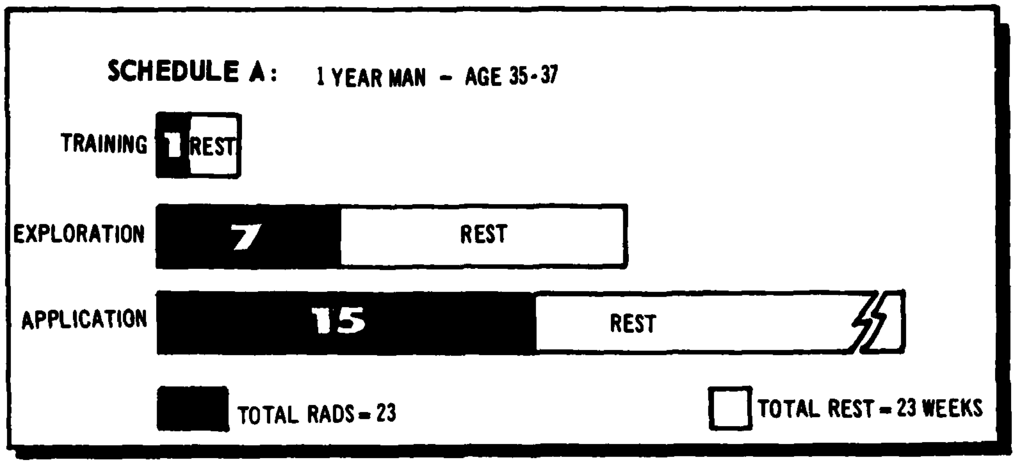


Figure 11-3.

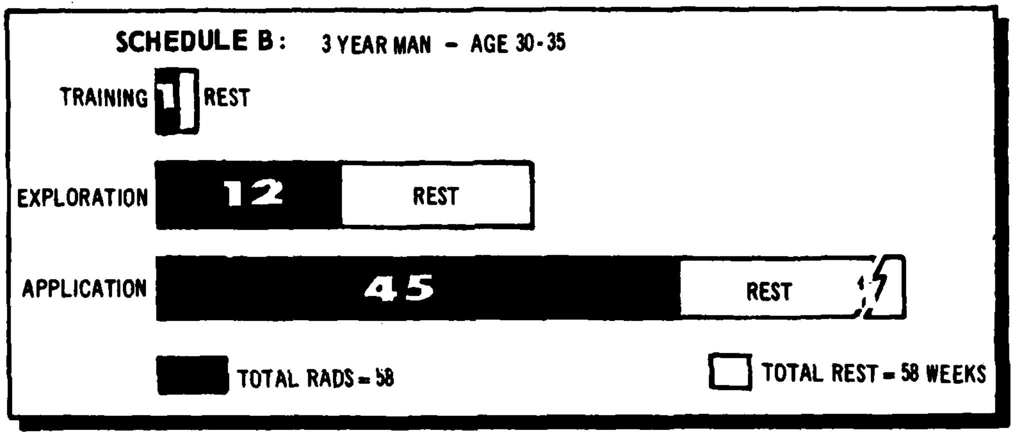


Figure 11-4.

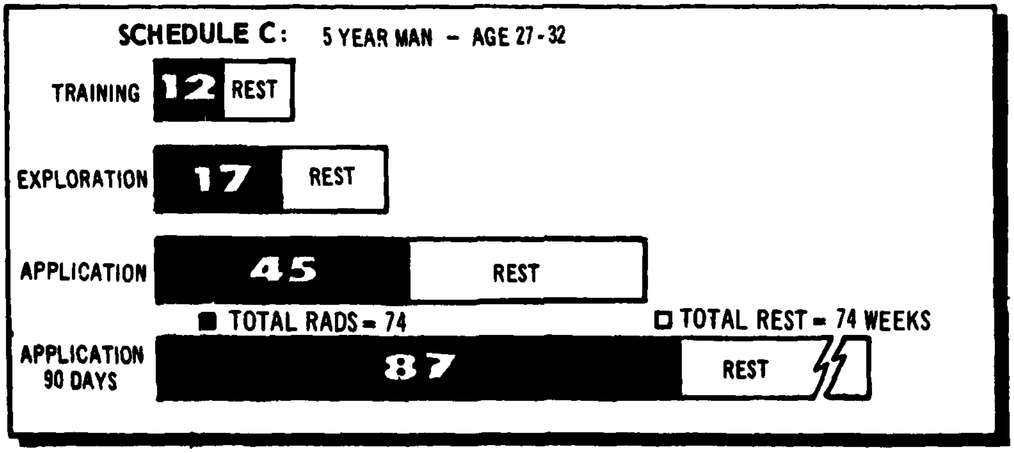


Figure 11-5.

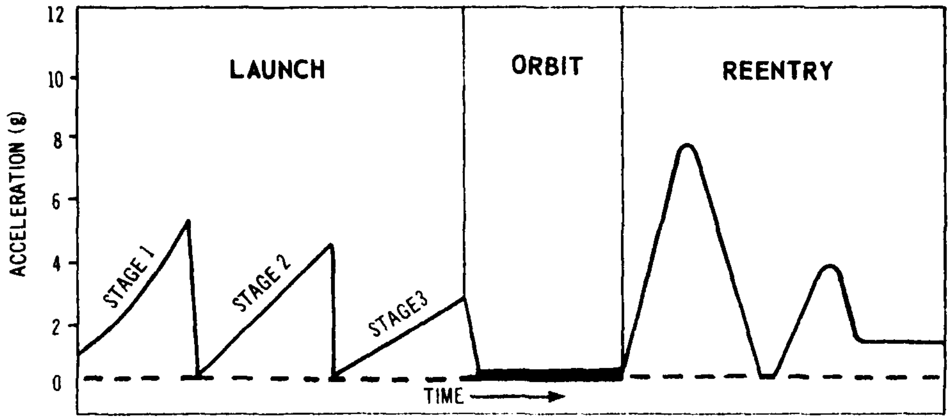


Figure 11-6.

which the g force is applied. When the acceleration is positive (from head to foot), the astronaut can expect an average tolerance of five g's for very short periods of time or two to three g's over longer periods. Negative g forces (foot to head) are less tolerable and a maximum of two to three g's is the tolerance level. Each of these tolerances would be marginal for space flight operations. Positive and negative g forces cause pooling of the blood in the extremities because the heart is unable to overcome the acceleration forces and complete the circulatory process. However, properly orienting the individual in the vehicle so that the g forces are transverse to the axis of the body—applied from front to back or back to front--can avoid these conditions. In this position, the g force is not acting on the long hydrostatic columns of blood that exert heavy pressures on the heart. In fact, even at high-transverse accelerations, the peak g load on the heart should be approximately one to two g's with proper care (see fig. 11-7). Therefore, our present manned systems place the crew member in a reclining position, with the legs and head raised slightly. With the acceleration forces acting in the transverse direction, the astronaut can withstand about 20 g's for short periods of time.

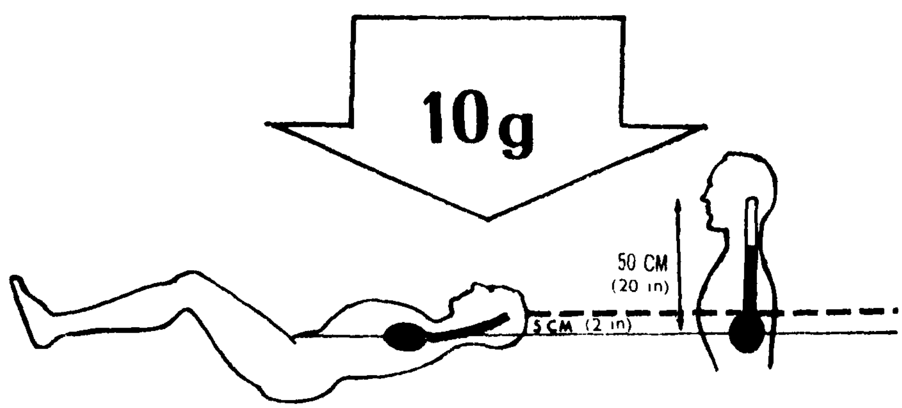


Figure 11-7. In the spacecraft, acceleration protection for the cardiovascular system is given by reclining on a couch so that the force is transverse. At 10g blood can reach the head with sufficient pressure to perfuse the brain. In the upright position the blood will not reach the head.

The arms, legs, and head present difficult problems under high-acceleration forces. For example, the astronaut cannot control the arms efficiently if g forces exceed about four g's. The hand and fingers have tolerance levels of about seven to nine g's. The astronaut's control and operation of the spacecraft would be limited or impossible if the astronaut used conventional aircraft controls. Therefore, engineers devised new controls and control locations so that the astronaut would have no difficulty operating the vehicle under high-acceleration forces. The astronaut can use devices such as side-stick controls, which keep the arm rigid but allow hand movement, for pitch, roll, and yaw of the space vehicle.

Vibration. A space launch and the rapid flight through successive layers of the atmosphere can expose an astronaut to rather severe vibrations from the rocket engines and buffeting in the atmosphere. Vibrations are side-to-side jostling motions and up and down bouncing. The astronaut is most sensitive to vibration frequencies of 4 to 10 cycles per second because the major internal organs have a natural resonance frequency in this range. When the organs become resonant with the vibrations of the space vehicle, severe pain, nausea, headaches, and dizziness are common symptoms. The organs begin to tear away from the mesentery holding them in place. Therefore, scientists and engineers must eliminate or dampen the vibrations in the frequency range of 4 to 10 cycles per second to make space flight safe for the astronauts. Under present power and flight schedules used in manned flight, we can control sufficiently the vibration problems. They produce no serious drawbacks to the program.

Noise. Noise has been a problem for people for a long time, but never before has the intensity been so great as the noise produced by space boosters. We know that people working in noisy environments develop a restricted deafness in the frequency range of the noise. It is necessary to place very rigorous protective controls on workers, including the use of ear plugs and ear muffs. The maximum tolerance to noise for people is about 140 to 150 decibels, at which level permanent damage to the hearing mechanism will occur if exposure continues for approximately one minute. Space boosters produce noise levels of 145 to 175 decibels. The solution to the noise problem involves using some of the natural physical conditions of space vehicles. For example, the capsule is placed on the end of the booster away from the rocket motors. The distance, which lessens the noise, reduces the decibel level in the capsule. Also, the materials in the vehicle, including the skin, structure, propellants, and oxidizers, as well as the equipment on board the spacecraft, help lessen the noise intensity. Finally, the space suit, which is worn during the operation of the first stage booster, has sound and vibration reduction materials throughout the suit and in the helmet. The manned space programs have employed these techniques successfully, and the program has kept the noise level in all vehicles well within the limits of safe operation.

Weightlessness. There were many anxieties concerning the effects of weightlessness on people operating in space. These ranged from fear of falling, nausea, and injury to the more complex physiological considerations, such as disruption of normal biological functions. The increase in our knowledge and experience have dispelled most of these anxieties. Although we have learned much about the effects of weightlessness on people, we still must proceed cautiously as we expand our capability. At the present time, we can divide the problems posed by weightlessness into two categories: psychological and physiological.

The pshychological problems are many and varied. Factors such as isolation, restraint, and confinement apparently have presented no particular problems in flights to date. However, as we anticipate longer flights, these three factors may present some serious problems. This was emphasized by Col Frank Borman on the Apollo 8 lunar orbiting mission, when he said, "We found several things we think need to be improved. I think that we need to concern ourselves with proper engineering for better body waste disposal systems. We have to provide showers on board. We have to have some sort of entertainment, television or canned tapes. We have to look into better foods. We have to realize that when you put a man on orbit for 60 days, or perhaps even longer, you have to pay more attention than we have in the past to the basic *creature comforts.*" The Skylab flights used such suggestions with good results.

Another area of concern is the sleep pattern for space flight. The normal physiological clocks of the astronauts were upset during the flight schedules of the early missions. For example, in low-earth orbit, the astronauts experienced approximately 15 day-night cycles in a 24-hour period. Also, when in weightless flight, the astronauts seem to need less sleep than they do in the 1-g environment of earth. As a result, the program planners have juggled the work-rest schedules in our missions to find the optimum regime. Because all missions vary so much, finding an optimum schedule has not been an easy task, and the program planners cannot make clearcut timetables. However, the longer duration Skylab missions have demonstrated the astronaut's adaptability to a new environment and within a short period, the astronauts established sleep patterns suitable for sufficient rest.

The psychological stresses facing astronauts in space flight have not been too great thus far. In the future, when flights of months and years become a reality, we may face serious psychological barriers. Much study is necessary to ensure safe space flights for long periods of time.

There are many physiological and adaptive changes that occur when an astronaut is exposed to a zero-g environment. This subject is one of the most controversial, and possibly the least understood area, related to manned spaceflight. Some feel that the problems involved will be the crux of the question concerning a person's ability to operate in the space environment. Others feel that a person's adaptive processes are so great that there will be no problem in placing astronauts in the weightless environment for extremely long periods of time.

We know that as people live in the presence of the earth's gravitational field, it produces certain effects within the body. There is relative displacement of organs, tissues, and fluids as a result of their different densities and varying orientations to the direction of the gravitational force. If there are any physiological changes resulting from weightlessness, they must be the result of the apparent removal of the gravitational force. What are the effects? Gravity is associated so strongly with the normal human physiological activities that the total effects of its removal are difficult to determine. We have experienced several thousand hours of weightless flight and have noted changes in several of the organs and organ systems, but these changes have not made the environment intolerable.

Cardiovascular deconditioning is probably the most significant problem. We have noted changes in blood volume, blood fluid shift, changes in the cellular ratios, and general deconditioning of the muscles and elastic properties of the blood vessels. These changes occur in the first 10 to 12 hours of exposure and are related directly to the weightless state of the blood and, as a result, the general reduction in work required to move blood throughout the body. There seems to be no particular problem adapting to the weightless environment, and it is possible that the system tends to reach a new steady state, at least in the short flights to date. The cardiovascular changes that occur seem to be within the tolerance limits of a normal individual for the periods involved. The major problem arises when the astronauts reenter the earth's atmosphere. They will experience high-deceleration forces and rapid return to the normal force of gravity on the cardiovascular system. It is possible they may experience orthostatic intolerance and other related stresses until the cardiovascular system can readapt to the gravity environment.

Muscle deconditioning and body mineral imbalance are other physiological disturbances that have appeared during weightlessness. There has been some discussion of possible psychophysiological disturbances related to long term weightlessness such as degradation of alertness and attention, vestibular function, and long-term loss of vestibular and proprioceptive sensory information. Although not seriously upsetting at this time, possibly these functions in long-term flights when combined with other psychological and physiological stresses may cause serious problems.

One of the earliest and most consistent manifestations of exposure to weightlessness is the now classical syndrome of "fullness of the head." This results from shifts in the body fluids and, in particular, the cephalad fluid in the head. The effect is something like hanging upside down. These fluid shifts occur early, and are responsible for several cardiovascular changes which are not well understood. It is quite possible, for instance, that they cause responses in the neurological and vestibular systems that result in the "space motion sickness syndrome." [The vestibular system includes the organs of the ear, which permit proper orientation and movement.] These same fluid

shifts may be the culprit responsible for the early diuresis, loss of effective circulating blood volume, and postflight problems in standing without dizziness or nausea. Other notable effects of weightlessness include a decrease in red blood cell mass, increased rates of calcium loss, and changes in body potassium and hormones affecting kidney function. A decrease in total body weight following spaceflights is a consistent finding and probably reflects losses in the lean body mass. Scientists have documented inflight changes in posture and increases in height as well.

In search for methods of overcoming the debilitating effects of weightlessness, scientists found that the symptoms were remarkably similar to those of immobile bed patients. These patients suffered blood volume changes, fluid shifts, orthostatic intolerance, weakened muscles, and calcium loss. The treatment for these ailments has been exercising the patients within their capabilities. Similarly, regular and vigorous preconditioning of astronauts and regular in-flight physical exercise have prevented the physiological changes from going beyond the tolerance limits.

The Skylab medical experiments form the backbone of our current understanding of the effects of prolonged exposure to weightlessness. That mission involved approximately 1½ man-years in orbit, with the longest duration being 84 days. Scientists scrutinized the astronauts closely before, during, and after the mission. Not only were there teams of doctors on the ground, but some of the astronauts were themselves MDs. Table 11-1 gives a gross breakout of the major areas of investigation.

Table 11-1.
Skylab Life Sciences Experiments

Area of Investigation	*No. of Experiments*
Nutrition and musculoskeletal function	4
Cardiovascular function	2
Hematology, immunology, and cytology	6
Neurophysiology	2
Behavioral effects	1
Pulmonary function and energy metabolism	2
Total	17

Despite the many questions still remaining, one clear result was the astronauts' ability to adapt successfully to this unnatural state of weightlessness. Now we are confident that an astronaut can tolerate missions of a year's duration, based on present knowledge. As we gather and analyze more data, it may be possible to increase this time significantly in the future.

Large, slowly rotating space stations that provide artificial gravity are a favorite with science fiction writers. These are certainly a real possibility, but there are some problems associated with this concept. Besides the effort and cost involved, scientists must address the coriolis forces. In oversimplified terms, coriolis is the apparent force you feel if you try to walk from the edge of a rotating disc toward the center. Stated another way, imagine yourself inside a giant tire that is rotating slowly. If you walk in the direction of the rotation, you will feel heavier; walk in the opposite direction, and you will feel lighter. Now imagine that we reduce the diameter of the tire approximately 20 feet. When you are standing still, the apparent gravity at your feet will be noticeably different from that at your head. These examples illustrate a few of the engineering difficulties associated with rotating space stations. Certainly, such problems are not impossible to solve, but they point out that artificial gravity is not a panacea.

LIFE SUPPORT EQUIPMENT

Without question, space presents an extremely hostile environment for people. In order to survive and operate a space vehicle, astronauts must have adequate protective devices to allow them to

remain in a homeostatic condition. Biomedical and ecological items comprise the life support equipment designed to given an astronaut protection during space flight.

Biomedical

In short or long-term exploration, we must be able to monitor the health of an astronaut and prescribe medication if the need arises. As evidenced on the Apollo missions, the transfer of germs from one astronaut to another in the close confines of our present space vehicles is rapid and unavoidable. Therefore, medical monitoring equipment is essential. Modern medical instrumentation allows monitoring of many physiological functions on a real time basis. For the purpose of immediate health monitoring, we are using a biomedical belt, or harness. This harness, equipped with the proper electrodes and receiver-transmitter units, relays the data on a variety of physiologial functions to the earth sensing stations via telemetering methods. These functions include the electrocardiogram, blood pressure, respiratory rate, electroencephalogram, and body temperature. Under the present program, scientists monitor only heart and respiratory rates in real time. They perform other medical test before a flight begins and when the crew returns to the earth. These tests analyze urine, feces, and blood samples. Possibly, the future will bring increased monitoring of the crew members on long term flights. This will be true when flights begin to carry astronauts to the distant reaches of our planetary system where there will be time delays in transmitting information back and forth between the space vehicle and the earth.

Should it be necessary to administer medication, the astronauts have rather complete medical kits on board. These small and compact kits (6 by 4.5 by 4 inches) contain motion sickness and pain suppression injectors, first aid ointment, eye drops, nasal sprays, compress bandages, adhesive bandages, thermometers, and a variety of antibiotic, nausea, stimulant, pain, decongestant, diarrhea, and sleeping pills. On the advice of medical personnel, the crewmembers can use the medicines in the kit to relieve most problems anticipated on present missions. Doctors can vary the medical kits as the mission varies. However, if a medical problem is beyond the capability of the kit and does require medical attention, NASA will terminate the mission as soon as possible.

Ecological

Ecology by definition is the study of the relationship of organisms to their environment. This means the study of all cyclic environmental factors related to the normal life cycle of living organisms, such as atmosphere, pressure, temperature, water, light, food, life expectancies, predators, waste disposal systems, and others. As we can imagine, the relationship of people to the environment is very complicated and becomes an even greater problem when an attempt is made to pack the normal ecology into a small space vehicle. We can show a very much simplified description of one cyclic pattern with the chemical equations for photosynthesis and respiration. Green plants, those containing the substance chlorophyl, can combine carbon dioxide (CO_2) and water (H_2O) to make glucose ($C_6H_{12}O_6$) and oxygen (O_2) in the presence of sunlight. Animals can metabolize glucose and oxygen to produce carbon dioxide and water. The cycle is complete. The cycle is graphically displayed as follows.

$$
\begin{array}{lccccccc}
\text{Plant photosynthesis:} & 6CO_2 & + & 6H_2O & \xrightarrow[\text{Chlorophyl}]{\text{Sunlight and}} & C_6H_{12}O_6 & + & 6O_2 \\
 & \uparrow & & \uparrow & & \downarrow & & \downarrow \\
\text{Animal Respiration:} & 6CO_2 & + & 6H_2O & \xleftarrow{\text{Metabolism}} & C_6H_{12}O_6 & + & 6O_2
\end{array}
$$

The balance of the many cycles that make up our closed ecological system on earth is much more complicated than shown with the sample equations.

The basis for the approach to life support equipment design is the analysis of human requirements on the one hand and the synthesis of the equipment to provide protection within the human tolerances on the other. To learn how scientists developed the environmental control systems, let us

first discuss the food, oxygen, and water requirements for people.

As we show in figure 11-8, in an open ecological system, a person has a requirement for approximately 22.7 pounds of oxygen, 4.7 pounds of water, and 1.3 pounds of food per day. At the same time, a person produces an equivalent weight of waste products, including 20.7 pounds of oxygen, 5.2 pounds of water, and 2.2 pounds of carbon dioxide. At present, it is impossible to carry these stored materials on space missions because of the weight involved. Scientists have made some progress in weight saving in the food program. Because of the mechanical problems of eating food during weightlessness in orbit, special preparation and packaging is necessary.

OXYGEN - WATER - FOOD EXCHANGE		
INPUT LBS/DAY		OUTPUT LBS/DAY
OXYGEN (O_2)	22.7	20.7 (O_2)
WATER (H_2O)	4.7	5.2 (H_2O)
		2.2 WATER VAPOR
		3.0 URINE
FOOD	1.3	2.2 CARBON DIOXIDE (CO_2)
(3000 KCAL/DAY)		0.6 SOLID WASTE
	28.7	28.7

Figure 11-8.

Our food system consists of dehydrated foods prepared by a process of freeze-drying. Then, processors package this water-free food in special plastic bags. The astronauts can rehydrate the food from the hot or cold water on board the spacecraft or, in the case of some items, by saliva in the mouth. This current food system has the advantages of being bacteria-free, lightweight, easily stored, and easily prepared. When rehydrated, the food is tasty and nutritious. However, it still has the disadvatange, inherent in an open ecological system, of being impractical from the viewpoint of long flights of months and years. Therefore, scientists must design some system of regenerating or growing food from the materials on board the spacecraft for future long-term missions.

The meals on Skylab were more palatable than those of previous manned programs. This was due primarily to advancements in technology that allowed the inclusion of frozen and canned foods.

The Skylab food system had to meet the rigid requirements and objectives of medical experiments that demand precise knowledge of nutrient intake. The food system provided the energy requirements of each individual astronaut based on body weight and age. The ration insured a daily intake between 750 and 850 milligrams of calcium, 1,500 to 1,700 milligrams of phosphorous, 3,000 to 6,000 miligrams of sodium, 300 to 400 milligrams of magnesium, and 90 to 125 grams of protein. Each astronaut maintained a constant level of intake of these controlled nutrients within 2 percent. Scientists baselined the diet to provide at least the dietary allowances of carbohydrates, minerals, vitamins, and fats recommended by the National Academy of Science.

Nutritionists stored approximately 950 kilograms (2,100 pounds) of food and accessories for all three manned visits aboard the workshop prior to launch. They secured the frozen food items in the five food freezers, and the other food items in lockers on the floor of the workshop at launch. Later, the SL-2 crew placed the food items in assigned locations during the activiation period. Crew

members taste tested the more than 70 food items provided in Skylab. The items aboard met the following criteria: food was a familiar kind; portions were processed to be prepared, served, and eaten in a familiar manner; and prepared food was satisfactory with regard to taste, aroma, shape, color texture, and temperature.

The meals on the Skylab menu included the following food types:

1. Dehydrated which includes ready-to-eat rehydratable foods with a moisture content reduced to less than three percent.

2. Intermediate moisture which includes precooked, thermally-stabilized, or fresh food with the moisture content reduced so that the final moisture content was approximately 10 to 20 percent.

3. Thermostabilized which includes precooked, thermally-stabilized, or fresh food with the temperatures reduced below 23°C (10°F) prior to launch to retard spoilage.

4. Frozen which includes precooked fresh food with temperature reduced below -40°C (-40°F) before launch to retard spoilage and maintained in freezers.

5. Beverages which include rehydratable drinks including black coffee, tea, cocoa, cocoa-flavored instant breakfast, grape drink, limeade, lemonade, orange drink, grapefruit drink, cherry drink, and apple drink.

Figure 11-9 shows an example of a typical three-day menu for one astronaut. Nutritionists provided one portable food tray for each crewmember. The food trays (fig. 11-10) had eight food cavities—four for large food cans and four for small cans. Three of the large cavities were heated. The astronauts set the timer and turned on the warmer. They used a removable tray lid when the food was heating. Each of the trays and lids had a color code. Magnets held the utensils to the tray until used. At the end of each meal, the astronauts filled the food trays for the next meal and set the heater timers for the appropriate intervals.

Each astronaut prepared foods that required rehydrating at mealtime through the use of the cold or hot water gun in the center of the food table. The food table pedestal housed the water chiller and the wardroom water heater. The manufacturers installed a plastic membrane inside each food can to prevent spillage in the zero-g environment. They fitted the membrane with a one-way spring-loaded valve for rehydratable food to permit the addition of water without leakage and to prevent the escape of contents during and after lid removal and mixing. The crew member slit the membrane with a knife, and the membrane retained the balance of the contents until the astronaut consumed them.

The manufacturers packaged rehydratable beverages individually in collapsed accordion-shaped beverage packs, which varied in length according to the type of beverage. These packs expanded in length with the addition of the cold or hot water to the beverage powder. The astronauts sat at the food table in special thigh restraints with their feet in portable foot restraints while eating.

The oxygen required by an astronaut requires investigation also. In an open ecological system, oxygen is a person's largest waste product. We breath in 22.7 pounds of oxygen per day, but, as figure 11-8 shows, we use only nine percent of the oxygen inhaled and throw away 20.7 pounds of oxygen. This is the largest portion of the total waste output by a person. Therefore, to make spaceflight economically feasible from a weight of storables standpoint, the gas portion of a person's system requires better treatment. To do this, the system processes the exhaled waste gases to remove the carbon dioxide and water vapor. Then the system returns the remaining oxygen to the astronaut.

On earlier space flights, all of which were of relatively short duration when compared with extended space travel, it has been simpler and more convenient to carry the entire oxygen supply in cylinders and to remove excess carbon dioxide with lithium hydroxide absorbers. Numerous and successful experiments during the last two decades have shown the practicability of oxygen replenishment by plants in a closed system, with the green plants obtaining their needed carbon dioxide from human metabolism. Therefore, research physiologists, engineers, and biologists are working to close all the various loops of an open ecological system. Figure 11-11 shows closed loops for oxygen and water.

This system does not account for the astronaut's food. However, a system can produce food by using algae or bacteria where these organisms will convert CO_2 and H_2O wastes into foods high in protein and carbohydrates. Also, these organisms will use (as they do on earth) the urine and fecal

Meal	Day 1	Day 2	Day 3
A	Scrambled Eggs Sausage Patties Strawberries Bread Jam Orange Juice Coffee	Cornflakes Chocolate Instant Breakfast Grape Drink Coffee	Scrambled Eggs Sausage Biscuit Cocoa Coffee
B	Chicken and Gravy Asparagus Peaches Biscuit Cocoa Lemonade	Cream of Potato Soup Chicken and Rice Pre-Buttered Roll Peaches Lemonade	Cream of Potato Soup Pork & Scalloped Potatoes Green Beans with Cheese Sauce Pears Grape Drink
C	Veal & Barbecue Sauce Mashed Potatoes Green Beans with Cheese Sauce Peach Ambrosia with Pecans Grapefruit Juice	Filet Mignon German Potato Salad Pears Biscuit Cocoa Grape Drink	Lobster Newberg Mashed Potatoes Asparagus Vanilla Wafers Vanilla Ice Cream Lemonade
	Snacks: Coffee x 2 Butterscotch Pudding	Snacks: Coffee x 2	Snacks: Dried Apricots Coffee x 2

Figure 11-9. Sample menu.

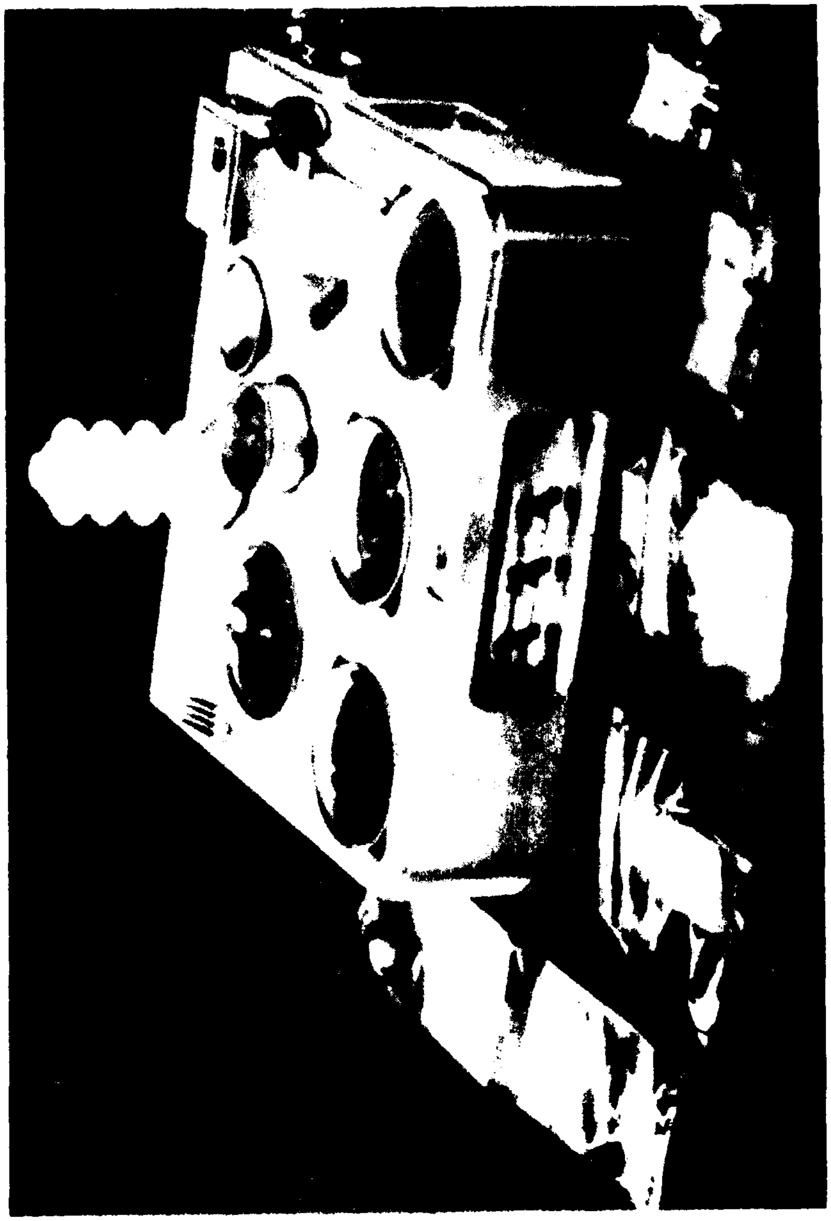


Figure 11-10. Food tray.

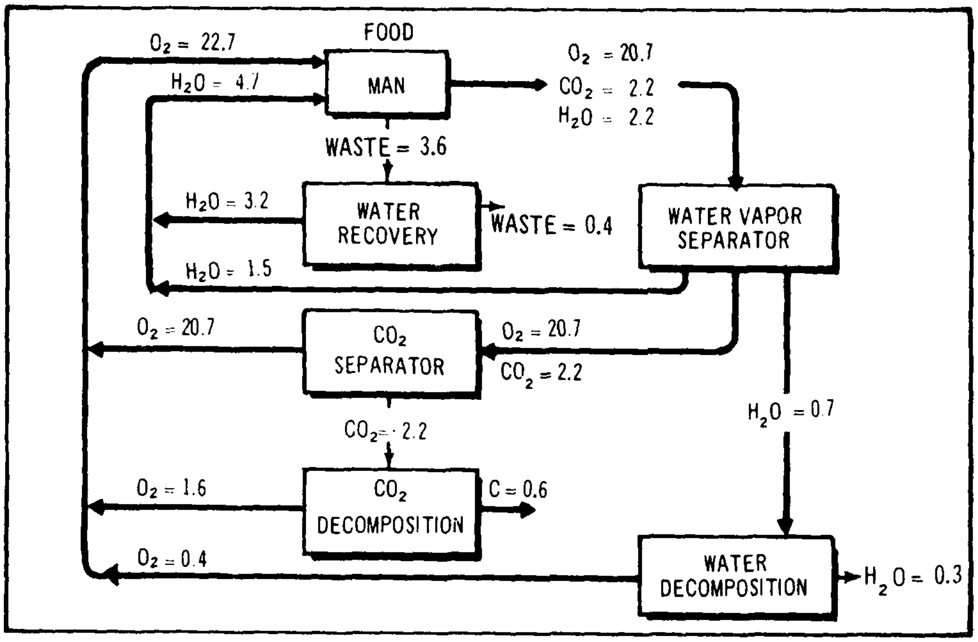


Figure 11-11. Closed gas and water system.

material of a person in their nutrient broth as plant food. This system will provide all the oxygen and food requirements for the astronauts and rid them of waste products. Eventually, scientists will design a completely closed and balanced ecological system that will include the astronauts as an essential part of the loop. The astronauts' waste products will be the food for the microorganisms, and the microorganisms' waste products will meet the food, water, and gas requirements of the astronaut. This is the system under which we are living on earth now.

In addition to the cabin environmental control systems, the astronauts must have environmental suits to protect them in the event of loss of cabin pressure and for extravehicular activity (EVA). Because the environmental factors do not change for the astronaut, the suit must provide the same protection as the cabin provided. Therefore, scientists have developed an undergarment that provides attachments for a communication belt, urine-collecting equipment, and the biomedical instrumentation belt. This suit is quite comfortable and worn during all activities in the space vehicle. Before the astronauts begin EVA, they switch to a similar underwear-type suit in which channels carry cool water that dissipates the body heat by sublimation of a secondary water source in the backpack. Over the undergarment, the astronauts wear a combination suit for pressure, micrometeorite, and radiation protection. The pressure portion of the suit provides 3.75 pounds per square inch pressure to all parts of the body. The helmet, in addition to the communication equipment, provides filters of various densities to protect the astronaut from the blinding glare of the sun. The multilayered external portion of the pressure suit provides protection against radiation and micrometeorites. This material is made of two inner layers of Beta cloth to intercept micrometeorites, three to four layers of insulating cloth, a layer of special metalized fabric, and another layer of Beta cloth. The material and design of the suit do not inhibit movement severely, while providing sufficient protection for activities outside the space vehicle.

BIBLIOGRAPHY

Colvin, Melvin and Oleg G. Gazenko. *Foundations of Space Biology and Medicine*. Washington D.C.: NASA Scientific and Technical Information Office, 1975.

Gerathewohl, Siegfried J. *Principles of Bioastronautics*. New York: Prentice-Hall, 1963.

Langley, Louis L. *Outline of Physiology*. New York: McGraw-Hill, 1961.

Morgan, Clifford T. *Human Engineering Guide to Equipment Design*. New York: McGraw-Hill, 1963.

Morgenthaler, George W. and Gerald E. Simonson. *Skylab Science Experiments: Proceedings of a Symposium of the 140th Meeting of the American Association for the Advancement of Science*. Tarzana, Calif.: AAS Publications Office, 1975.

National Aeronautics and Space Administration. *Compendium of Human Responses to the Aerospace Environment*. August 19 21, 1970. NASA Contract Report 1205. Washington D.C.: National Aeronautics and Space Administration, November 1968.

National Aeronautics and Space Administration. *Fifth Symposium on the Role of the Vestibular Organs in Space Exploration*. National Aeronautics and Space Administration Special Paper 314. Washington D.C.: National Aeronautics and Space Administration, 1970.

Parker, James F., Jr and Vita R. West. *Bioastronautics Data Book*. 2d ed. National Aeronautics and Space Administration Special Paper 3006. Washington D.C.: National Aeronautics and Space Administration, 1973.

Robinson, Douglas H. *The Dangerous Sky*. Seattle: University of Washington Press, 1973.

Chapter 12

MILITARY SPACE SYSTEMS

Historically, military operations have evolved successively through the areas of land, sea surface, underwater, air, and space. Each new area of operation has been important to total military capability, and the optimum military posture requires proper cross support and cooperation. As technologies advance, the emphasis on the various regimes and the nature of their interlocking varies, but no new operational medium has, in itself, provided a panacea for solving all military problems. The Department of Defense (DOD) has been analyzing constantly the military contributions that they can obtain from space operations. The DOD is trying to strike a proper balance between developments in the newly-opened regime of space and military efforts in the older regimes of land, sea, and air. According to AFM 1-6, *Aerospace Basic Doctrine, Military Space Doctrine,* "To achieve their full potential, space systems and operations must be integrated fully within existing military forces to become part of the total force structure."

There is a limitation on the DOD in developing space systems that could threaten the security of other nations. Under the terms of the outer space treaty, which went into force on 10 October 1967, the United States agreed not to place weapons of mass destruction in outer space or in orbit around the earth. The treaty bans military maneuvers on celestial bodies. As a result, the DOD is using the medium of space for warning, surveillance, communication, navigation, and weather systems to assist in maintaining an effective deterrent posture.

SYSTEMS AND PROGRAMS

In this section we will describe some of the space programs undertaken with the active participation of the DOD and Air Force. These programs range from those of immediate peacetime value, such as communication satellites, to those with potential military application.

Space Warning System

One of the best examples of using space systems to augment surface systems would be to provide an early warning of an intercontinental ballistic missile (ICBM) attack. Already, operational early warning satellites are in geostationary orbits to warn of an ICBM or Fractional Orbiting Bombardment System (FOBS) attack and to warn of a submarine-launched ballistic missile (SLBM) attack on the continental United States. We have one satellite stationed over the Eastern Hemisphere and two over the Western Hemisphere.

The eastern satellite would provide the first warning of an ICBM raid on the United States by the Soviet Union or the Peoples Republic of China. The Ballistic Missile Early Warning System (BMEWS) would verify this warning. The capability to correlate data from BMEWS, satellites, and other sources provides a highly credible warning of attack. The western satellites provide the first warning of SLBM launches against the United States with verification provided by FSPS-77, FPS-85, and FPS-115 radars.

Military Space Communications Systems

Space communications systems can augment ground-based command and control systems. Basically, a command and control system is a composite of equipment, skills, and techniques capable of performing the clearly-defined function of enabling a commander to exercise continuous control of the forces and weapons in all situations. It provides the commander with the information needed to make operational decisions and with a means for disseminating those decisions.

As advancements are made in technology, the Air Force is using this technology to develop a communications satellite system that is more jam resistant, provides two-way communications among mobile units, tactical aircraft, the national command authorities, and the strategic forces. The complexity, range, and speed of military systems continue to increase. As the need for positive communications becomes critical, space-based systems will provide an ever increasing amount of communications.

There are five military space communication programs that the Space Division (SD) of the Air Force Systems Command is managing. These include the Defense Satellite Communications System (DSCS), the Fleet Satellite Communications System (FLTSATCOM), the Air Force Satellite Communications System (AFSATCOM), the Satellite Data System (SDS), and Milstar.

Defense Satellite Communications System. Phase II of the DSCS is a second generation program that has replaced the successful initial DSCS (see fig. 12-1). The space portion of DSCS II consists of spin-stabilized advanced communications satellites in synchronous orbit around the earth. Being significantly more powerful than the previous satellites, the phase II spacecraft provide high-capacity, super high-frequency (SHF) secure voice and data links for the Worldwide Military Command and Control System (WWMCCS). They support terminal deployments for contingencies; restoration of disrupted service overseas; presidential travel; global connectivity for the Diplomatic Telecommunications Service; and transmission to the continental United States of some surveillance, intelligence, and early warning data.

The DSCS II operational system consists of four active satellites and two spares orbiting the earth. Each satellite contains propulsion systems for orbit reposition to support contingency operations. Ground command can steer the two dish-shaped antennas on DSCS II. The antennas can concentrate their electronic beams on small areas of the earth's surface for intensified coverage to link small, portable ground stations into the communication system.

In Phase III, the antenna design allows the user to switch between fixed earth-coverage antennas and multiple beam antennas (see fig. 12-2). The latter will provide an earth-coverage beam as well as electrical steerable area and narrow-coverage beams. In addition, a steerable transmit dish antenna will provide a spot beam with increased radiated power for users with small receivers. In this way the communication system will tailor the communication beams to suit the needs of different size user terminals almost anywhere on the surface of the earth. The DSCS III satellite has a cross-dipole transmit antenna and a flatback cavity spiral receiver antenna as a part of the Air Force Satellite Communication (AFSATCOM) program link with the strategic alert forces. The first Phase III satellite was launched in October 1982, but the complete Phase III system is still being developed.

Presently, the Satellite Control Facility (SCF) at Sunnyvale, California, handles the communication capabilities and positioning of DSCS II through a worldwide network of SCF stations. The planners and designers of the DSCS III system will design it so that selected Defense Communications Agency (DCA) managed satellite configuration control elements (SCCE) will have the ability to control both the satellite's communication capabilities and its position as well. The current operation plan for DSCS III requires that the SCCE assume the primary role in controlling the satellite communications systems with the AFSCF continuing to perform the functions of controlling the other satellite systems and maintaining the proper satellite orbit as in DSCS II. This will increase the flexibility of DSCS by providing a more direct response to communications system user requirements and by providing DSCS III with backup capability for the SCF, should it ever be needed.

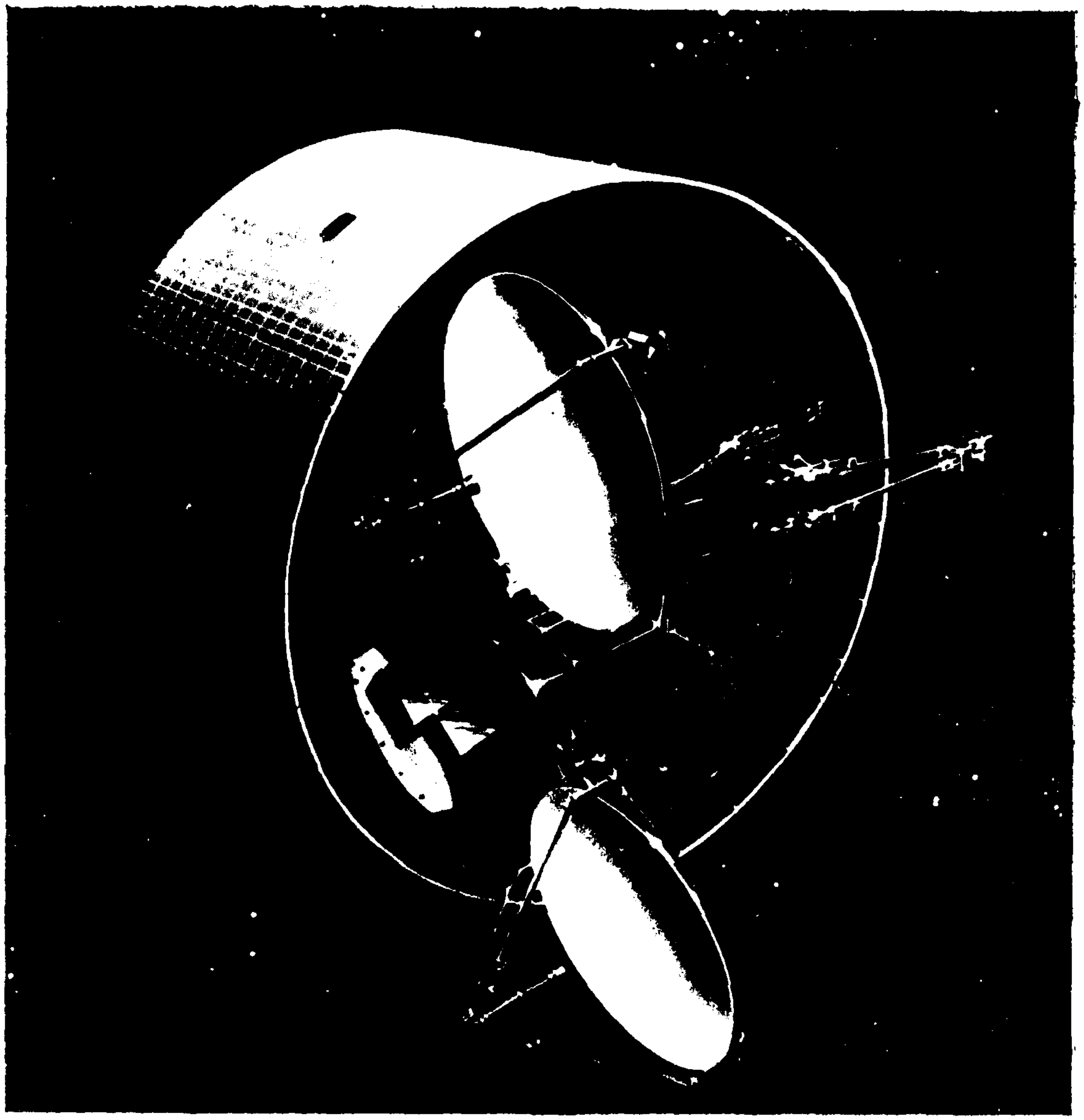


Figure 12-1. Defense Satellite Communications System Phase II

The Air Force Titan III launch system can launch all DSCS satellites. On 30 October 1982, a single Titan 34D with an inertial upper stage successfully placed into geosynchronous orbit a DSCS II and DSCS III spacecraft. The designers of the DSCS III plan for on-orbit delivery of it by a space shuttle. They schedule the first launch of two DSCS III satellites on the shuttle for 1985.

The Air Force System Command's Space Division (SD), Los Angeles Air Force Station, California, manages the design, development, production, and launch of the DSCS spacecraft. The DSCS II spacecraft are 9 feet in diameter, 13 feet tall, and weigh 1,200 pounds. They are spin stabilized at 150 revolutions per minute. The DSCS III spacecraft are approximately 9 feet in diameter, 7 feet tall, and weigh 1,900 pounds. They are three-axis stabilized.

Figure 12-2. Defense Satellite Communications System Phase III.

Fleet Satellite Communications System. The Fleet Satellite Communications System (FLTSATCOM) is a major and essential step in modernizing Navy communications (see fig. 12-3). It helps to relieve the Navy of its almost total dependence on high-frequency transmission for beyond-the-horizon communication and adds needed capabilities not possible at high frequency, such as antijam fleet broadcast. Operating at ultrahigh frequency, FLTSATCOM allows relatively low-cost terminals with simple antennas for use on highly mobile platforms. Unlike the DSCS, FLTSATCOM has a relatively small capacity because of its much lower operating frequency. The FLTSATCOM system provides a satellite communication system for high-priority communication requirements of the Navy and Air Force that encompasses almost the entire world. This system supports other DOD needs as well.

The space segment consists of four satellites in geosynchronous equatorial orbit. Each satellite has 23 communication channels in the ultrahigh and superhigh frequency bands. The Navy has exclusive use of 10 channels for communication among its land, sea, and air forces worldwide. The Air Force uses 12 others as part of its (AFSATCOM) system for command and control of nuclear capable forces. The satellite system has one 500 kilohertz channel allotted to the national command authority. The satellite's hexagonal body is 8 feet in diameter and 50 inches high. The main

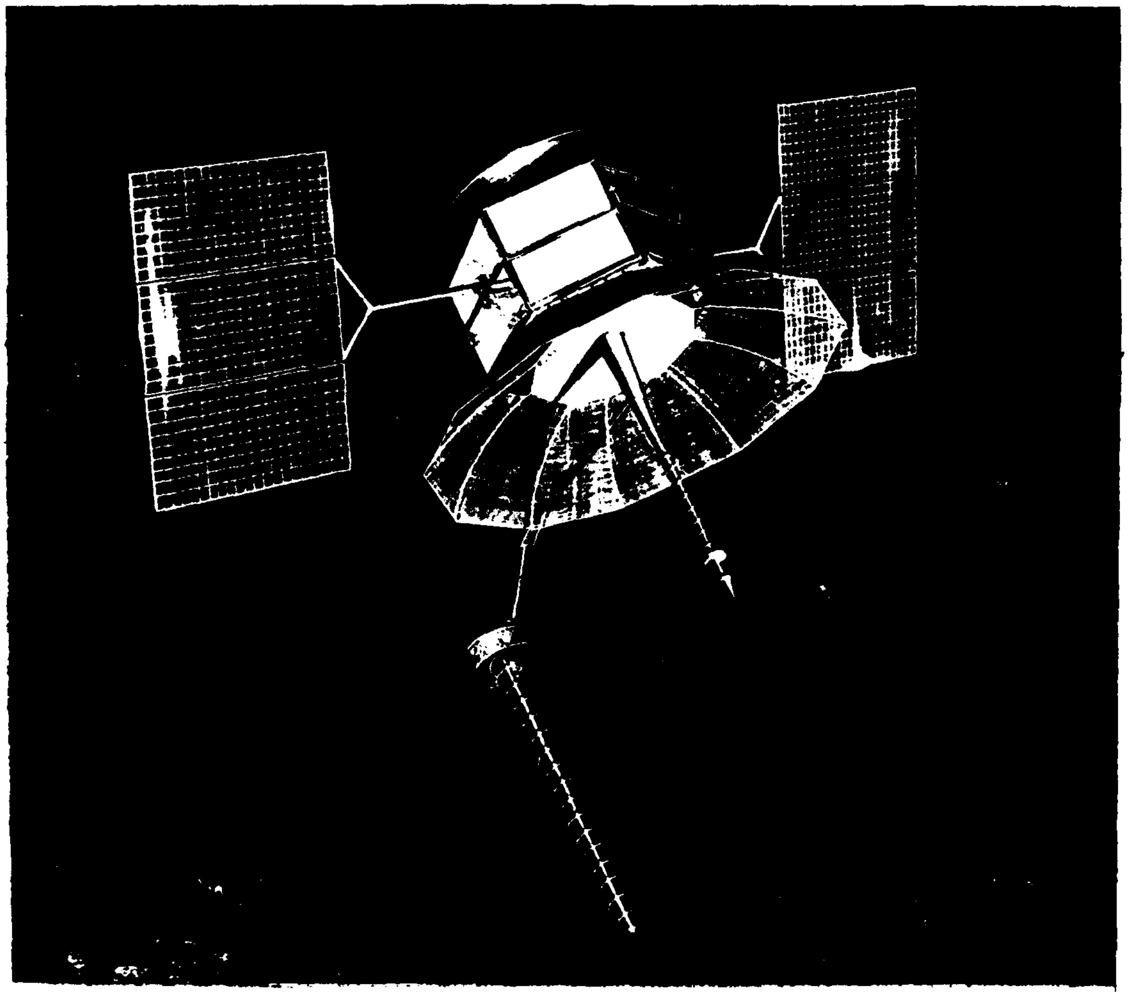


Figure 12-3. Fleet Satellite Communications System.

parabolic transmit antenna is 16 feet in diameter with an 80 inch solid center surrounded by wire mesh screen. The designers of the satellite mounted a 13.5 foot stepped helical receiving antenna, 13 inches in diameter at the base, outside the edge of the antenna dish. The satellite and its Apogee Kick motor with propellant weigh approximately 4,100 pounds.

The ground segment of the FLTSATCOM system consists of communications links among designated and mobile users, including most US Navy ships and selected Air Force and Navy aircraft, as well as global ground stations.

The Atlas-Centaur serves as the launch rocket for each FLTSATCOM spacecraft. After the Atlas-Centaur places each vehicle in an elliptical orbit, an onboard solid-fuel Apogee Kick motor lofts the satellite into geosynchronous orbit. A body-fixed momentum wheel interacts with monopropellant hydrazine thrusters to provide three-axis attitude stabilization and point the satellite's antennas at the earth's center. Hydrazine jets allow the user to change the satellite's location in orbit.

The Naval Electronics Systems Command has overall program management responsibility. The program office at Space Division (SD) manages the acquisition of the space segment of the program.

Air Force Satellite Communications System. A general war-survivable satellite communications system for command, control, and communications (C^3) of our nuclear capable forces must withstand massive physical and jamming attacks in the execution of its mission. This is the goal of the AFSATCOM program. The AFSATCOM-1, which has a modest antijam capability, consists of Air Force transponders installed on host satellites. AFSATCOM transponders on the Satellite Data System (SDS) are providing north polar coverage. Transponders on the FLTSATCOM satellites will provide equatorial coverage. Transponders installed on additional satellites in the SDS and FLTSATCOM provide proliferation of coverage to increase survivability. The Air Force is planning for AFSATCOM II, which will emphasize a major upgrade in antijam capability and improvement in satellite physical survivability.

AFSATCOM will employ short, low-speed messages for force execution, force report back, and force redirection. The use of such teletype messages coupled with suitable antijam techniques will permit relatively simple ultrahigh frequency low-power terminals aboard our operational vehicles to reliably and securely communicate by satellites that large land-based jammers stress.

Satellite Data System. The SDS Program is a multipurpose communications satellite program that provides polar coverage for command and control of the strategic forces. Additionally, it provides a high-data date communications link between the remote tracking station of Thule, Greenland, and the Satellite Control Facility at Sunnyvale, California.

Navigation Systems

Navigation is essential to military commands for precise delivery of weapons on designated targets, deployment of troops, and scheduled rendezvous of units at designated objectives. Increases in the military's mobility and in the sophistication and accuracy of their weapons have created the requirement to improve position-fixing and navigation capabilities. This and the worldwide distribution of military operations requires a navigation system with all-weather global coverage. At present, there are many navigation aids of various types that provide position-fixing capabilities over great sections of the earth for a large number of users.

Doppler shift influence. A major breakthrough in navigation procedures came with the first Soviet Sputnik in 1957 when scientists noted that a graphical frequency plot of the "beeps" received from the satellite formed a characteristic curve of Doppler shift. As a satellite approaches or recedes from a tracking station, the transmitted radio frequency appears to change. The amount of this Doppler shift, or change, is dependent upon the relative position of the tracking station with respect

to the path of the satellite. If we make an accurate measure of the Doppler shift and know the precise orbit of the satellite, we can calculate the precise location of the receiving station. A system based on this fact could be passive, operable in all weather, and global in coverage. It could provide good navigational accuracy. Its major drawback could be the length of tracking time required of a receiving station.

This concept was the basis for the Navy's Transit system and the follow-on Navy Navigation Satellite system, an all-weather navigation system that has been in use since 1964. It has enabled naval fleet units equipped to use the system to accurately establish their positions anywhere on the seas.

NAVSTAR Global Positioning System. The NAVSTAR Global Positioning System (GPS), known as NAVSTAR, is a space-based radio navigation network that will satisfy the precise positioning and navigation needs of all the military services (see fig. 12-4). In the fully operational system, satellites circling the globe every 12 hours will beam continuous navigation signals to earth. With proper equipment, a user can process the signals and determine position within tens of feet, velocity within a fraction of a mile per hour, and time within a millionth of a second. The fact that the satellite employs an atomic spaceborne clock with an approximate drift rate of 10^3 seconds is a unique feature of the system. If a user has a clock that is in synchronization with the satellite clocks, the user can measure the time difference between transmission and reception.

Figure 12-4. NAVSTAR Global Positioning System.

To receive this information a NAVSTAR user only needs to push a few buttons. A user's set will select automatically the four most favorably located satellites, lock onto their navigation signals, and compute the user's position, velocity, and time. Engineers are developing receiving sets for integration with aircraft, land vehicles, and ships. A lightweight backpack unit is under production and test for ground troops as well. Possible applications of the NAVSTAR include precision all-weather weapon delivery; enroute navigation for space, air, land, and sea; aircraft runway approach; photomapping; and range instrumentation and safety operations.

The NAVSTAR satellites are three-axis stabilized in a 10,900 nautical mile (20,178 kilometer) circular orbit with an orbit period of 12 hours. A fully operational system will position at least 18 operational satellites equally spaced in six orbital planes to provide global coverage. Each NAVSTAR satellite transmits three L-band, pseudo-random noise-coded signals, one S-band, and one ultrahigh frequency for spacecraft-to-spacecraft data relay.

The three general classes of user equipment are single channel, two channel, and five channel. This user equipment includes antennas, receivers, signal processors, flexible modular interfaces, and control/display units.

Five widely separated monitor stations track passively all satellites in view and accumulate ranging data from the navigation signals. The personnel at a NAVSTAR master control station receive this information for use in satellite orbit determination and systematic error elimination. The master control station communicates with the satellites through three ground antennas. Through these links, the control station updates the satellites with information so users receive optimum navigation data. Space Command acts as executive agent for the Department of Defense in managing the NAVSTAR GPS program.

Defense Meteorological Satellite Program

Timely knowledge of weather conditions is of extreme importance in the planning and execution of military field operations. Real-time night and day observations of current weather conditions provide the field commander with greater flexibility in the use of resources for imminent or ongoing military operations. The military has established firmly the importance of meteorological data from satellites in the effective and efficient conduct of military operations, and new applications continue to appear as the scope and quality of meteorological satellite data improve.

Although there are other meteorological satellites in use by the civilian community, the Defense Meteorological Satellite Program (DMSP) satellites are designed to meet unique military requirements for weather information (see fig. 12-5). During the Vietnam era, commanders used DMSP data gathered over Southeast Asia to plan daily air, sea, and ground operations.

Through the DMSP satellites, military weather forecasters detect and observe developing cloud patterns and follow existing weather systems. The data help identify severe weather such as thunderstorms, hurricanes, and typhoons. Visible and infrared imagery are used to form three-dimensional cloud-plural analyses of various weather conditions. An important feature of this imagery is its near constant resolution across the 1,600 nautical mile wide data swath. Although the primary mission of DMSP satellites is gathering weather data for military uses, its information is actually a national resource. Data gathered by the satellites are made available to the civilian community through the Commerce Department's National Oceanic and Atmospheric Administration (NOAA).

The Block 5D-2 satellite, launched in December 1982, weighs approximately 1,660 pounds including 400 pounds of sensor payload. It is 21 feet long with solar array deployed. The four major sections of the spacecraft are (1) a precision mounting platform for sensors and other equipment that requires precise alignment; (2) an equipment support module that encloses a bulk of the electronics; (3) the spent upper-stage rocket motor and supporting ascent phase, reaction control equipment; and (4) a solar array.

The sensor payload performs many missions for the DMSP system. The Operational Linescan

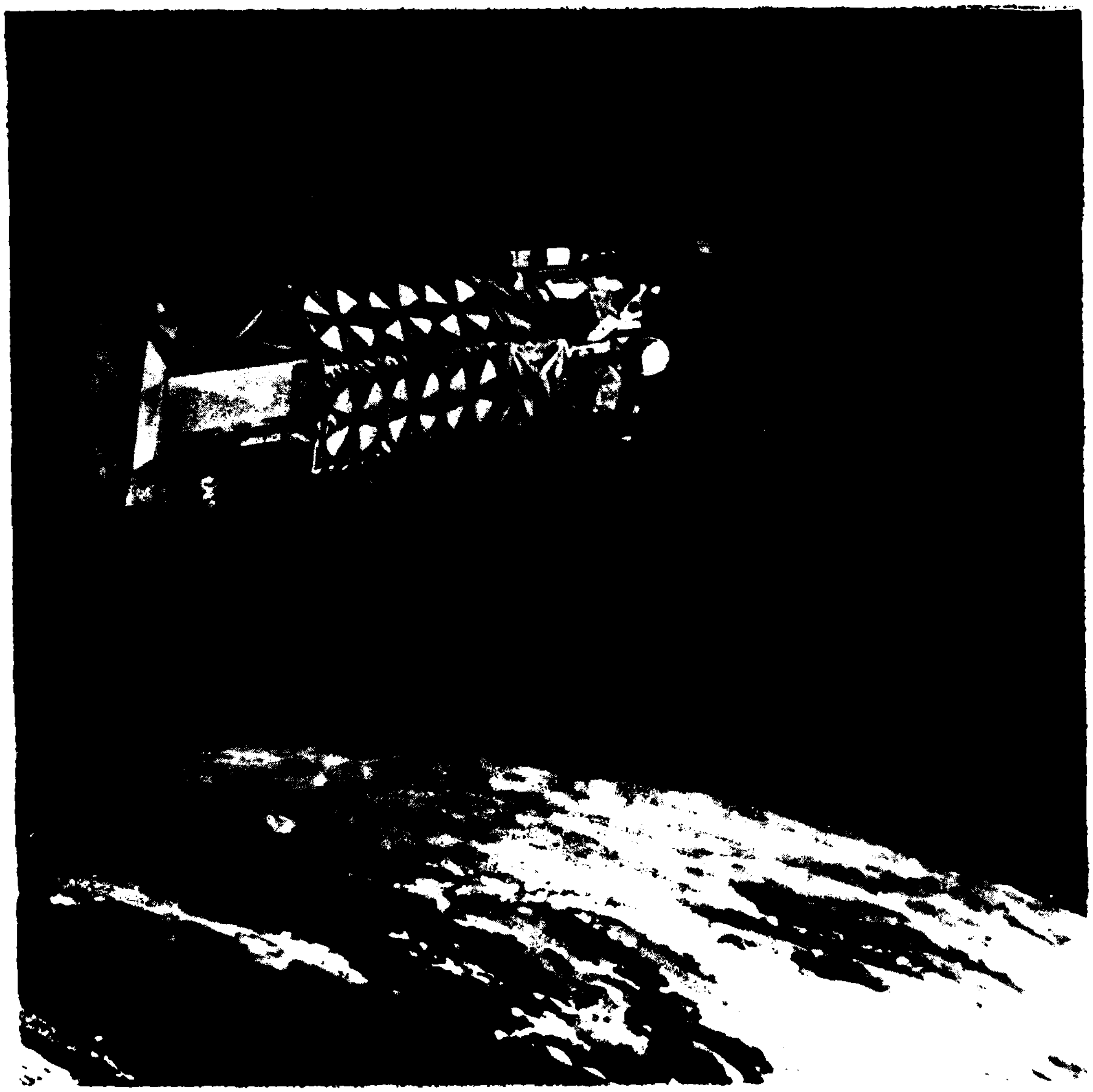


Figure 12-5. Defense Meteorological Satellite Program.

System (OLS) is the primary sensor on board the spacecraft providing visual and infrared imagery. This OLS uses a linescanning radiometer to take visual and infrared (day and night) imagery in both 0.3 nautical mile and 1.5 nautical mile resolution. The imagery that the OLS produces is near constant resolution across scan. The military uses these data to analyze cloud patterns in support of a wide range of military requirements.

The fourth 5D-2 satellite will contain an additional imaging system known as the mission sensor system, microwave imager (SSM/MI). The sensor is a passive microwave radiometer that detects and images microwave energy that the atmosphere and surface of the earth emit. These measurements will provide military meteorologists information about ocean surface wind speed, age and coverage of ice, areas and intensity of precipitation, amount of water and clouds, and moisture of soil.

DMSP satellites carry sensors that make temperature (moisture) measurements. The infrared

temperature/moisture sounder measures infrared radiation emitted from different heights within the atmosphere, allowing forecasters to plot curves of temperature and water vapor versus altitude. The satellite uses a microwave temperature sounder to measure microwave radiation emitted from different heights within the atmosphere. This instrument allows forecasters to plot curves of temperature versus altitude even over cloudy regions of the globe.

The DMSP satellite uses another group of three sensors to make measurements of the earth's ionosphere. It uses the first one, known as a precipitating electron and proton spectrometer, to count electrons and protons at different energies spiralling down (precipitating) along the magnetic field lines. The military, among others, uses these data to forecast accurately the location and intensity of the aurora and to aid radar operations and long-range ground communications in the Northern Hemisphere. A second sensor measures energy in the high-energy gamma and X-ray portion of the spectrum, which are responsible for the aurora. A third ionospheric sensor, the ionospheric plasma monitor, measures electron and ion densities and temperatures at spacecraft altitude. Ionospheric weather modules use data from the sensor to predict the state of the ionosphere as it affects communication and navigation.

Launched on Atlas boosters, two of these Block 5D-2 satellites orbit at any one time. They orbit at an approximate altitude of 450 nautical miles in near-polar, sun-synchronous orbits. Each scans a 1,600 mile wide area and can image the entire earth every 12 hours.

DMSP satellites record data throughout their orbits on tape recorders and play the data back to ground stations. The satellites are also capable of providing weather data on real-time basis to the Air Weather Service and Navy terminals, which provide military commanders in the field with photographic-quality prints of cloud cover four times a day.

The Air Force Space Division, Los Angeles Air Force Station, California, manages the design, development, production, and launch of these spacecraft. The 1000 Satellite Operations Group of the Space Command operates the DMSP system command and control functions.

Space Test Program

In May 1965 the director of defense research and engineering established the DOD Space Test Program (STP), (originally called the Space Experiments Support program), with the Air Force as executive agency. Initially, the director chartered the program to support developmental and preoperational payload. However, it later increased the scope of the program to include the research payloads supported previously by the Office of Aerospace Research. Since its first launch in 1967, the STP has flown successfully more than 95 payloads on 21 different spacecraft missions. The program has support from all 3 services as well as 30 different laboratories and program offices. They are responsible for work on ballistic missile defense, communications, geodesy, navigation, atmospheric research, space object identification, solar physics, spacecraft subsystem development, radar calibration, and particle and radiation studies.

SPACE SURVEILLANCE

When the Soviet Union launched SPUTNIK I on 4 October 1957, the United States became aware that it had little capability to detect, track, or identify man-made objects in space. This realization led to an intensive effort to create and maintain a system of sensors to keep track of satellites. This effort continues today. The mission of the US space surveillance system is to detect, track, and maintain surveillance of all military payloads in space. The space surveillance network of ground-based sensor systems accomplishes this through data collection. The network uses this data to classify and identify all detected objects, maintain an accurate and current catalog of space objects, and provide data reports to military and civilian agencies and to the scientific community. These data reports consist of orbital and signature data on space objects. They provide new space launch detection and tracking, foreign satellite function identification, satellite maneuver identification, collision avoidance information, data on satellite transiting of specific geographical areas, space object reentry impact points, warning of attack on US space assets, targeting information for the

US antisatellite system, and successful and unsuccessful attack verification.

Command and control and sensor components comprise the space surveillance network. The Space Defense Operations Center (SPADOC) accomplishes the command and control. It is located in the NORAD Cheyenne Mountain Complex (NCMC) near Colorado Springs, Colorado. There is an alternate center at Eglin AFB, Florida. The NORAD Space Computational Center (NSCC) receives the data collected by the sensors for processing to create data reports (see fig. 12-6). The network collects and transmits the data to the NCMC for processing and dissemination. Although the types of sensors differ, their space surveillance missions remain essentially the same. Sensors function in either dedicated, collateral, or contributing roles.

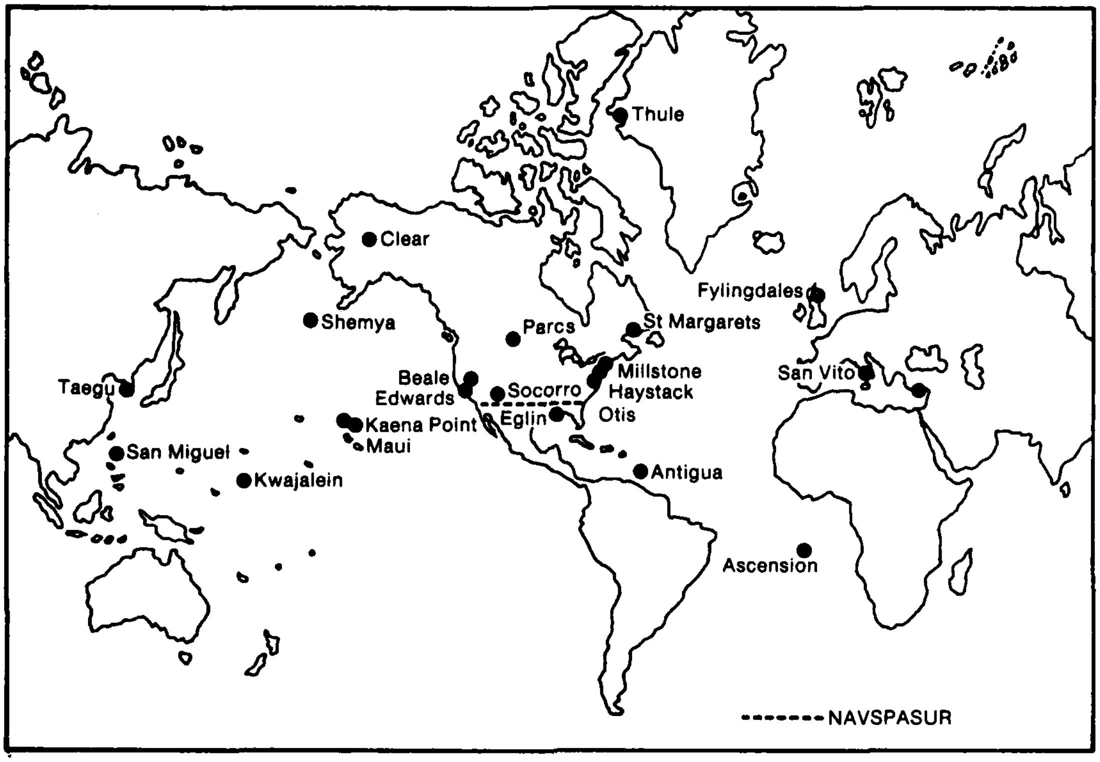


Figure 12-6. Sensor Surveillance Network.

Dedicated Sensors

The role of the dedicated sensor is extremely important in space surveillance. This system consists of the Baker-Nunn cameras (optical); ground-based optical deep-space surveillance (GEODSS, electro-optical) system; Maui Optical Tracking and Identification Facility (MOTIF, electronic-optical); the Navy Space Surveillance System (NAVSPASUR, radar); and the Global Positioning System-10 (GPS-10, radar).

Optical sensors. One type of dedicated sensor is the optical sensor. The three basic optical sensors are the Baker-Nunn cameras, the GEODSS, and the MOTIF.

The Air Force owns two Baker-Nunn cameras. Civilians operate and maintain them under contract. The Canadian military operates and maintains one additional camera. The Baker-Nunn system is a film camera attached to a large deep-space telescope. The system collects data on

satellites ranging in altitude from 3,000 nautical miles to geosynchronous (22,000 nautical miles) and beyond. The film data collection sites at Edwards AFB, California; St. Margarets, Newfoundland, Canada; and San Vito, Italy, transmit the data to the NCMC in two to four hours after collection. Film processing and analysis requires most of this time.

The Air Force owns three GEODSS sites with a fourth and fifth site programmed. A civilian contractor operates and maintains this system. The developers of the GEODSS designed it to collect data from 3,000 nautical miles to beyond 22,000 nautical miles as does the Baker-Nunn camera. The GEODSS sites, unlike the Baker-Nunn, provide the data in near-real time to the NCMC. GEODSS uses a low-light-level TV camera, computers, and large telescopes. GEODSS is more sensitive than the Baker-Nunn in that it can track smaller and dimmer objects, and collect space object identification (SOI) signature data. The site locations are Socorro, New Mexico; Maui, Hawaii; Taegu, South Korea; and Diego Garcia (Indian Ocean). The Air Force has not selected a site for the programmed fifth site.

The MOTIF site is an Air Force owned site, colocated with GEODSS at Maui, Hawaii. MOTIF possesses technical capabilities similar to GEODSS. The site is contractor operated and maintained.

Radars. The Navy Space Surveillance (NAVSPASUR) system is a network of sensors, known as radiometric interferometers, stretching from Georgia to California. Three transmitters, six receiver sites, and a control center make up the NAVSPASUR system. The newest addition to the space surveillance network is the AN/GPS-10 radar located at San Miguel, Republic of the Philippines. The radar is Air Force operated and contractor maintained. The GPS-10 radar is the only Air Force radar dedicated to the space surveillance mission.

Collateral Sensors

Some sensors have the capability of performing more than one task at a time and have space surveillance missions as secondary missions. The military considers sensors having this capability as part of the space surveillance network. SPADOC tasks these sensors for observational data as long as this assignment does not interfere with their primary mission. These sensors are responsible to NORAD for their primary mission and are called collateral sensors. The nominal range of the sensors is in the near-earth arena up to approximately 3,000 nautical miles. The exception is the FPS-85 at Eglin AFB, Florida, that can operate at higher altitudes. The sensor types and locations include the AN/FPS-85, phased array, at Eglin AFB, Florida; Cobra Dane at Shimya, Alaska; BMEWS at Clear, Alaska, Thule, Greenland, and Flyingdales, England; PARCS, phased array, at Concrete, North Dakota; AN/FPS-79 at Princlik, Turkey; and the Pave Paws, phased array, at Otis ANGB, Massachusetts, and Beale AFB, California.

Contributing Sensors

Contributing sensors are those not under NORAD's operational control. However, they do provide observational data on satellites on a contributing basis. The Air Force Eastern Test Range under the Eastern Space and Missile Center at Patrick AFB, Florida, and the Air Force Western Test Range under the Western Space and Missile Center at Vandenberg AFB, California, receive data from and contribute data to the space surveillance network. The Massachusetts Institute of Technology Lincoln Laboratory complex at Westford, Massachusetts, contributes data as well as the Kiernan Reentry Measurement System (KREMS) located at Kwajalein Atoll. The sensor names and site locations include Kaena Point, Hawaii; ALTAIR and ALCOR sensors at Kwajalein Atoll; the Millstone and Haystack sensors at Westford, Massachusetts; and others on Antigua Island and Ascension Island.

System Improvements

The baseline space surveillance system is undergoing improvement. Areas being improved are

survivability; command, control, and communications; identification capability; coverage; tracking accuracy; and processing. The improvements include additions and replacements, upgrades to the sensor and communications systems, and a major upgrade of the processing and sensor-tasking systems.

Several radars are due for upgrades. The BMEWS radar upgrade will provide smaller range resolution and better tracking accuracy, attack assessment, and count capability. PAVE PAWS upgrades will track smaller payloads associated with MIRVs and will get a more accurate count for attack assessment purposes. There will be an upgrade to NAVSPASUR. It will include receiver transmitter modernizations that will increase the range to above 18,000 kilometers. C-band upgrades to Antigua and Ascension will enhance the search capabilities.

Some other improvements to the system include an addition of the Compensated Imaging System (CIS) to the telescope at the ARPA Main Optical System (AMOS); Improved Radar Calibration Systems (IRCS) at the three BMEWS radars, the FPS-79, GPS-10, and FPS-85; Charge Coupled Device (CCD) cameras in the GEODSS system for improved search rate, sensitivity, and reliability; state-of-the-art computer replacements; improved communications terminals at each missile-warning site; implementation of SPADOC Phase IV; and addition of secure voice terminals at all missile-warning and space-surveillance sites. In the future the Air Force plans to deploy space-based space surveillance systems to enhance the survivability and endurability of the space surveillance network.

BIBLIOGRAPHY

Henry, Richard, Maj Gen. "Space from a Fighter Pilot's Point of View." *Astronautics and Aeronautics*. (April 1976):33.

Miller, Barry. "US Moves to Upgrade Missile Warning." *Aviation Week and Space Technology*. (2 December 1974):16.

"New Weather Satellite Series Planned." *Aviation Week and Space Technology*. (15 July 1974):16.

Nichols, David A., Capt. "The Defense Meteorological Satellite Program." *Optical Engineering* 14, no. 4. (July-August 1975): 273.

Paparozzi, Frank, Lt Col and Capt Neal Anderson. "DOD's Space Test Program." *Astronautics and Aeronautics*. (June 1974):42.

Slay, Alton D., Lt Gen. "New Aircraft and Missiles." *Supplement to the Air Force Policy Letter for Commanders*. (May 1976):10.

Smith, Douglas, Maj and Capt William Criss. "NAVSTAR Global Positioning System." *Astronautics and Aeronautics*. (April 1976):26.

Stelling, Henry B. Jr., Maj Gen. "NASA-DOD-USAF and the Space Shuttle." *Supplement to the Air Force Policy Letter for Commanders*. (November 1975):7.

US Air Force. AFM 1-6. *Military Space Doctrine*. October 1983.

Chapter 13

SPACE TRANSPORTATION SYSTEM

The Space Transportation System (STS) is a national asset that provides routine access to space for both civil and defense users. Elements of the STS include the Space Shuttle, upper stages, Spacelab, launch and landing facilities, simulation and training facilities, and mission control facilities. The STS is a reusable system capable of deploying a wide variety of scientific and applications satellites (see fig. 13-1). Since it can carry payloads weighing up to 29,500 kilograms (65,000 pounds), it will ultimately replace most of the expendable launch vehicles currently in use. NASA can use it to retrieve satellites from earth orbit, service or repair and then redeploy them, or bring them back to earth for refurbishment and reuse. Scientists and technicians can use it to conduct experiments in earth orbit. It is an effective means for the United States to use current and future capabilities in space.

The Space Shuttle, which is the heart of the STS, includes two reusable solid-propellant rocket boosters and a reusable orbiter resembling a delta-winged airplane mounted piggyback on a large expendable liquid propellant tank (see fig. 13-2). The solid-propellant rockets and three liquid-propellant rockets on the orbiter launch the system to an altitude of approximately 44 kilometers (7.4 miles). At that point, the solid-propellant rockets separate from the system and land in the ocean for recovery, refurbishment, and reuse. The orbiter (see fig. 13-3). continues the flight with the liquid-propellant tank until main engine cutoff. Then, the engine is jettisoned to reenter the atmosphere and fall into the ocean, and the orbiter fires the engines of its orbital maneuvering system for a short period to gain power for insertion into earth orbit. It can remain in orbit with its crew and payload for a period ranging from 7 to 30 days and then return to earth and land like an airplane (see fig. 13-4).

This chapter examines the capabilities of the Space Shuttle and describes the major components of the system. They include the orbiter, the external tank, and the solid-propellant rocket boosters. It includes a brief discussion of launch facilities, ground support systems, and management of the Space Shuttle program.

SYSTEM MANAGEMENT AND CAPABILITIES

The National Aeronautics and Space Administration (NASA) is responsible for overall management of the STS. The Department of Defense (DOD) is responsible for representing national security interests in the STS and, therefore, is participating as a partner in its development, acquisition, and operations. The DOD has designated the Air Force as the sole point of contact with NASA for all commitments affecting the STS and its use in matters regarding national security space operations. The broad policies and principles that govern the relationship between the DOD and NASA relative to the STS are defined in a NASA/DOD memorandum entitled *Memorandum of Understanding on Management and Operations of the Space Transportation System.* It further delineates their roles and responsibilities in its development, acquisition, and operation.

Within NASA, the Space Shuttle program office is responsible for detailed assignment of responsibilities, basic performance requirements, control of major milestones, and funding allocations to the various NASA field centers.

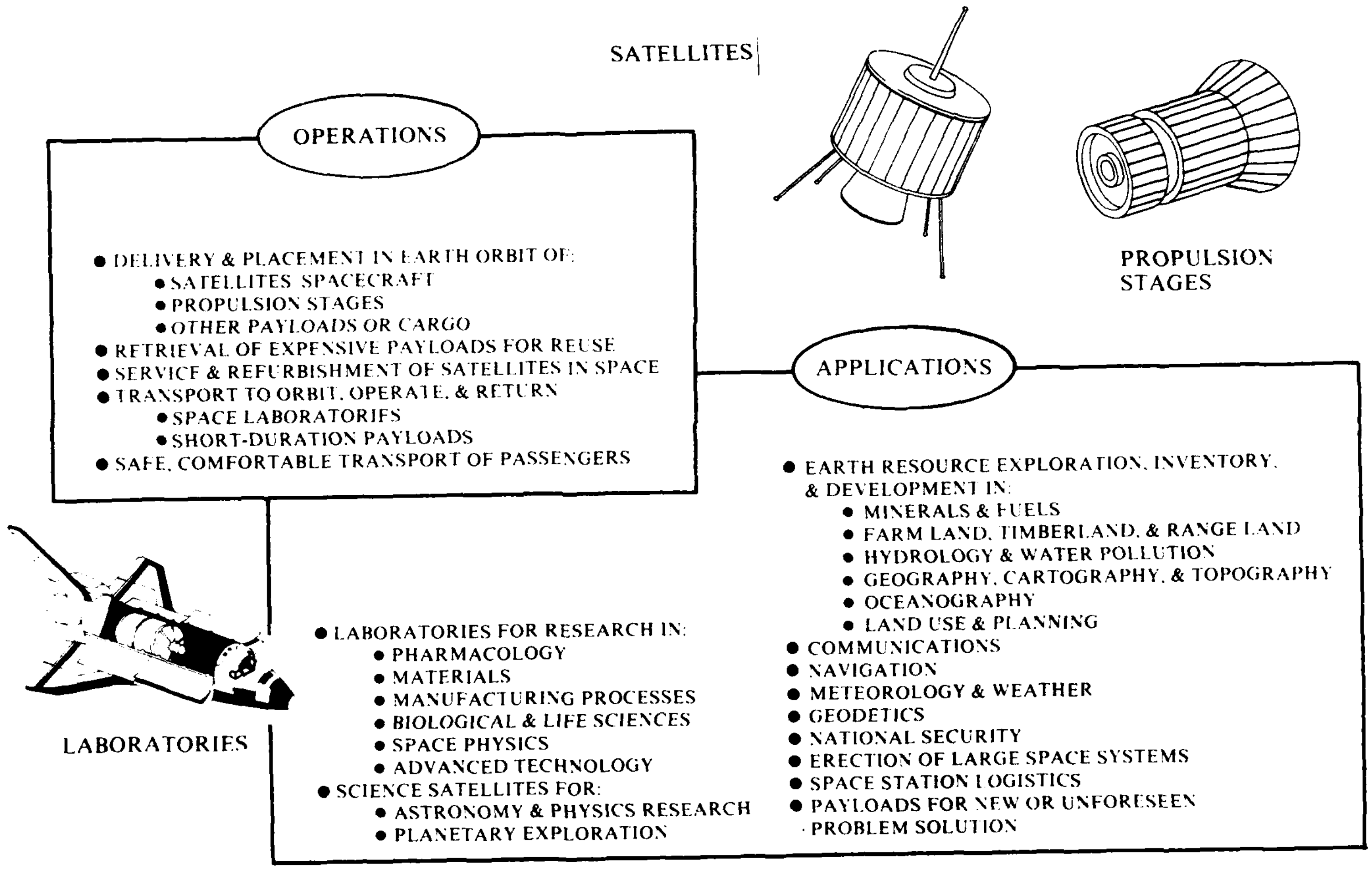
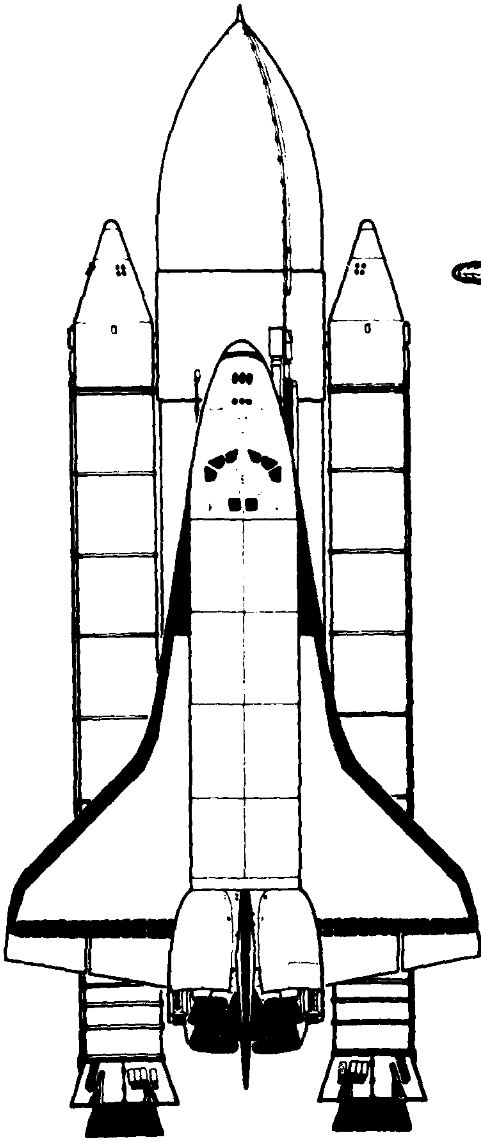
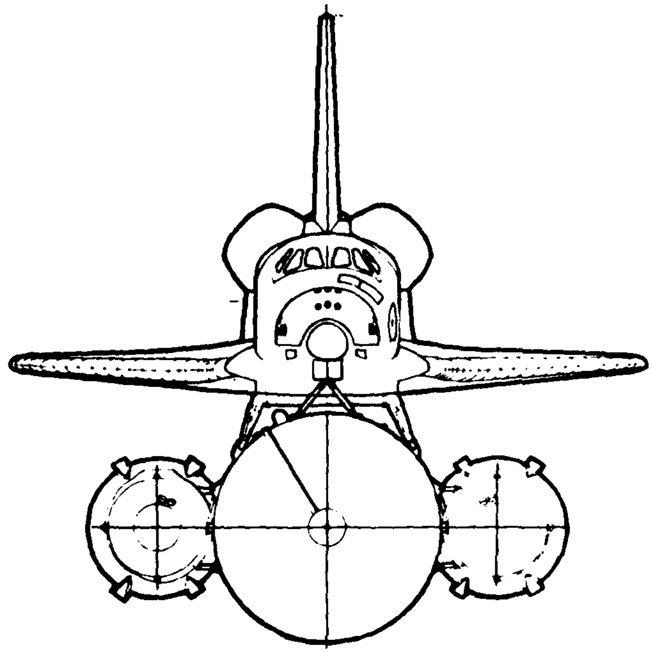


Figure 13-1. Potential missions of Space Transportation System.





ORBITER SPECIFICATIONS
Length: 121 feet
Width: 79 feet
Cargo Bay: 60 feet by 15 feet
Weight: 75 tons

EXTERNAL TANK
Height: 154 feet
Weight: 1.6 million pounds (full)

SOLID ROCKET BOOSTERS (SRBs)
Height: 149 feet
Weight: 1.3 million pounds each

Typical mission length: seven to thirty days
Typical crew: two to seven people
Height of orbit: 135-320 nautical miles (most missions)
Speed in orbit: 17,550 mph (at 150 miles)
Payload capacity: 65,000 pounds
Launch and landing sites: Kennedy Space Center, Fla.
Vandenberg Air Force Base, Calif.
Principal tracking station: White Sands, N.M.
Mission Control: Johnson Space Center, Texas

Figure 13-2. Space Shuttle—facts and figures.

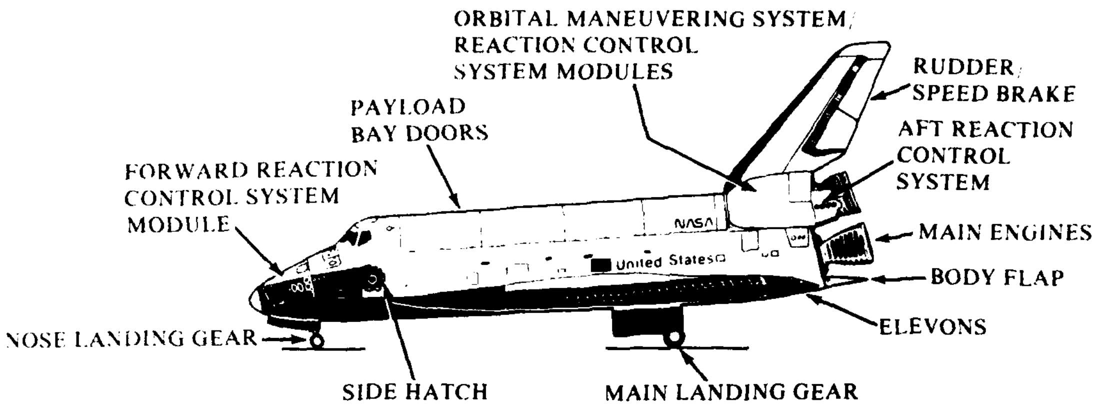


Figure 13-3. Orbiter.

The Johnson Space Center, Houston, Texas, is responsible for program control, systems engineering and integration, and definition of elements that interact with other elements, such as total configuration and combined aerodynamic loads. The center is also responsible for development, production, and delivery of the orbiter. The Kennedy Space Center, Florida, is responsible for design of launch and recovery facilities.

The Air Force Space Shuttle Program Office, Space Division, Los Angeles Air Force Station, California, is responsible for the design, development, engineering, and overall management of the Vandenberg AFB, California, shuttle launch and landing site, and the inertial upper stage to be used by the DOD. The office is responsible for DOD interface with NASA on all uses of the shuttle including payload integration, flight assignments, and security.

The Air Force Space Command is responsible for the operation of the STS through the Consolidated Space Operations Center (CSOC) located at Falcon Air Force Station, Colorado Springs, Colorado. This facility, expected to be completed by 1990, will be the center of all DOD shuttle flight operations.

The versatility of the shuttle provides a number of advantages in space exploration, but the primary benefit will be reduced costs of building and launching satellites. First, it provides the capability to retrieve payloads for repair and reuse when they fail or show critical malfunctions. Second, moderate acceleration and low noise inside the payload bay will allow the designers to create simpler and less rugged designs for future satellites. Third, the shuttle, by itself or in combination with an advanced upper stage, permits retrieval of satellites for refurbishment and reuse when they reach the end of their service life. Fourth, it can carry replacement modules to failing satellites in orbit and eliminate the need to recover entire satellites. Fifth, NASA can use it to perform dedicated missions, test subsystems, or demonstrate new technology.

Satellites with missions requiring altitudes, inclinations, or trajectories beyond the capability of the shuttle will require use of an upper stage. The two basic upper stage systems are the spinning solid upper stages (SSUS-D and SSUS-A) and the inertial upper stage (IUS). NASA will use the spinning, solid, upper stages to place up to 3,800 pounds in geosynchronous orbit, and the IUS to place 5,000 pounds or more into geosynchronous orbit or planetary orbit (see fig. 13-2).

When the shuttle becomes fully operational, it will carry into space virtually all of the civilian and military payloads of the United States and many payloads from the international community, including private industry, universities, and research organizations.

SHUTTLE COMPONENTS

The Space Shuttle consists of the orbiter and the solid-propellant rocket boosters that can be

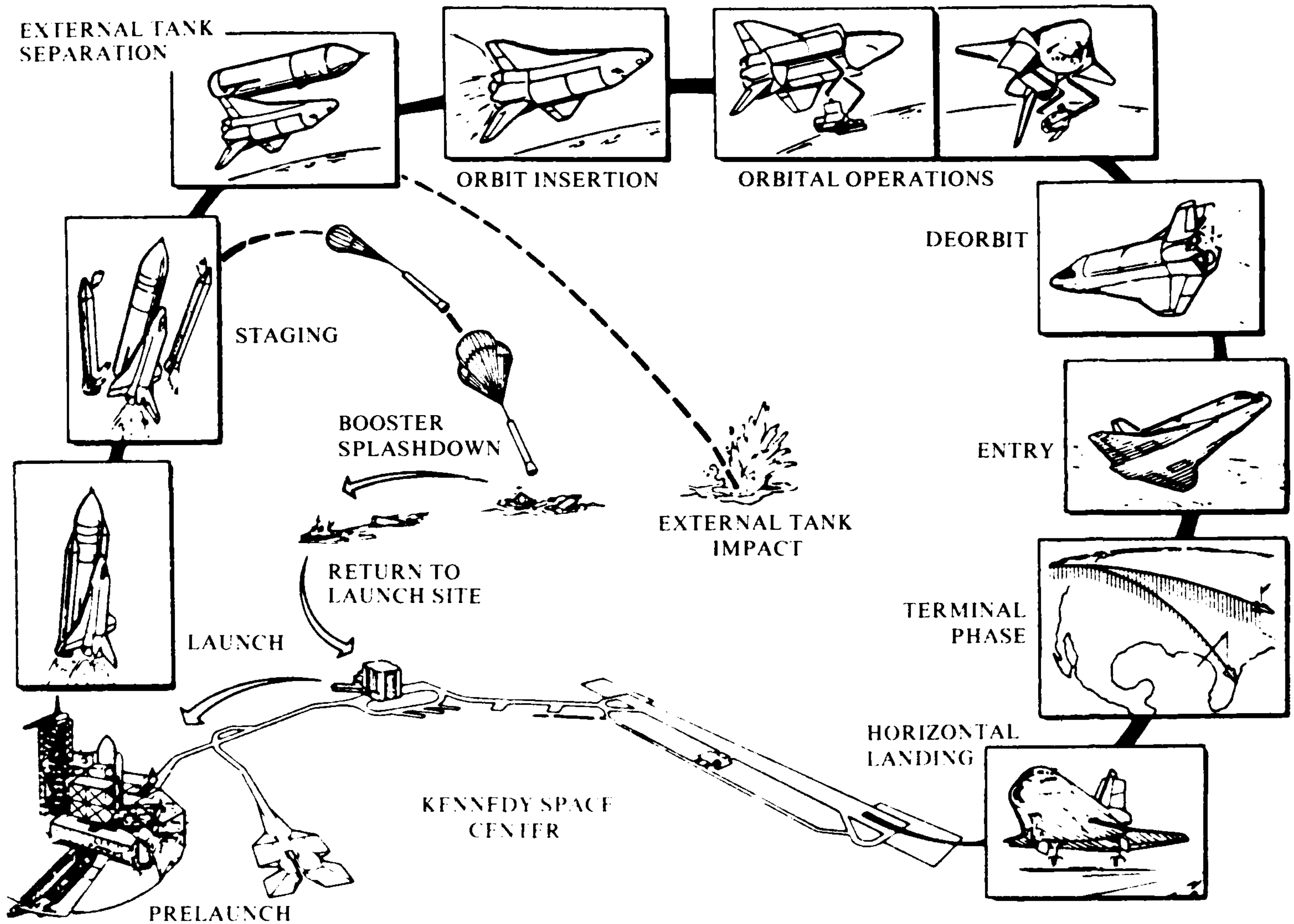


Figure 13-4 Typical mission profile

reused for future space launches. However, the designers of the system have designed the external tank, another component, for a single space mission. The design of the shuttle subsystems places much emphasis on ease of refurbishment and handling between space usage. Other spacecraft have used some of the technology that the shuttle uses. However, some technological advances are designed for exclusive shuttle use. The shuttle represents state-of-the-art technology for space vehicles.

ORBITER

The orbiter resembles an airplane and performs like one during the last minutes of flight. However, it is far more complex than the most sophisticated aircraft. It contains 49 engines, 25 antennas, 5 computers, separate controls for flying in space and in the air, as well as electric power generators that produce drinking water. Its dimensions include a length of 37 meters (121 feet), a wing span of 24 meters (120 feet by 79 feet), and a weight of approximately 75,000 kilograms (165,000 pounds) without a crew and a payload. It's cargo bay, 18.3 meters (60 feet) long and 4.6 meters (60 feet x 15 feet) in diameter, can deliver single or mixed payloads of 29,500 kilograms (65,000 pounds) to an orbit of 240 kilometers (150 miles) altitude. It can place smaller payloads in a maximum orbit of 1,110 kilometers (690 miles), and it can return payloads of 14,515 kilograms (32,000 pounds) to earth. Eventually it will carry a crew of three astronauts (pilot, copilot, and mission specialist) and one to four scientists or technicians to manage the payloads. The main sections are the forward fuselage containing the airtight crew module, the cargo-carrying midfuselage, the rear fuselage containing the engine thrust structure, wings that house the main landing gear, and the vertical tail.

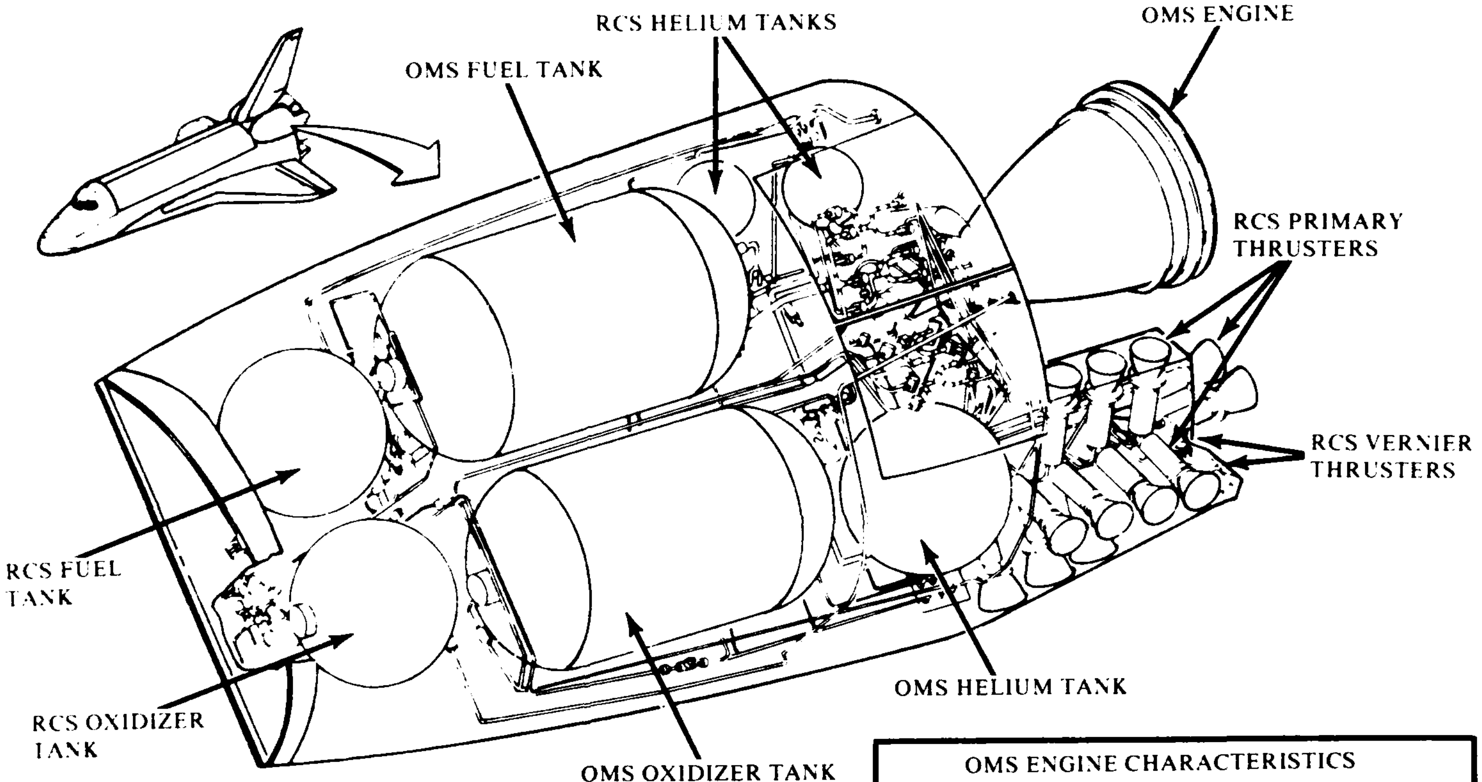
Reusable surface insulation covers the airframe, made primariy of aluminum, to protect the crew from the heat produced during reentry. It uses a special tilelike insulation that reflects heat so effectively that one side can be red hot and the other side can be cool enough to hold in bare hands. The two types of insulation that cover the top and sides of the orbiter are blocks of silica fiber with a glossy coating and flexible sheets of nylon felt coated with silicone. The tiles, approximately 2.5 centimeters thick and 20 centimeters square (1 inch by 7⅞ inches square), protect the aluminum surfaces from temperatures up to 650 degrees centigrade and the flexible insulation from temperatures up to 370 degrees centigrade. The coating gives the upper part of the craft a white color and optical properties that reflect solar radiation.

Similar tiles with a different coating protect the bottom of the spacecraft and the leading edge of the tail from temperatures up to 1,260 degrees centigrade. Unlike the coating on the upper part of the craft, the high-temperature coating gives the underside a glossy black appearance. Since the nose and leading edges of the wings reach the highest temperatures, reinforced carbon-carbon covers them for protection against temperatures up to 1,650 degrees centigrade. This is a carbon cloth impregnated with extra carbon and heat-treated with silicon carbide.

The three main engines used on the orbiter are the most advanced rocket engines ever built and the first engines designed for repeated use. Mounted in a triangular pattern on the rear fuselage, they can swivel 10.5 degrees either up or town and 8.5 degrees from side to side to change the direction of their thrust. They assist the solid-propellant rocket boosters during launch and carry the craft to its final orbiting position.

The Orbital Maneuvering Subsystem engines (see fig. 13-5) in external pods to the left and right of the upper main engine produce 6,000 pounds of thrust each. They accelerate the craft to orbital velocity after the main engines shut down and the external tank drops away. They supply energy to change orbits, rendezvous with other spacecraft, and return to earth. The spacecraft can use the engines separately or together and can swivel them plus or minus eight degrees to control direction.

Small rocket engines called reaction control thrusters (see fig. 13-6) in the nose and near the tail provide attitude control in space for changes in velocity during the final phases of rendezvous and docking or for orbital corrections. These engines and aerodynamic control surfaces on the craft control the attitude of the orbiter during reentry into the atmosphere and at high altitude. In the nose are 14 primary reaction control engines, each of which produces 875 pounds of thrust, and two vernier engines with 25 pounds of thrust for small corrections. There are 12 primary and 2 vernier engines nestled in each pod in addition to the maneuvering engine.



OMS ΔV CAPABILITY	1000 FT/SEC (65,000 LB PAYLOAD)
USEABLE OMS PROPELLANT	23,876 LB TOTAL
	14,866 LB N_2O_4
	9010 LB MMH

OMS ENGINE CHARACTERISTICS

THRUST	6000 LB (VACUUM)
SPECIFIC IMPULSE	313 SEC
CHAMBER PRESSURE	125 PSIA
MIXTURE RATIO	1.65
GIMBAL CAPABILITY	± 6° PITCH ± 7° YAW

Figure 13-5. Orbital Maneuver Subsystem.

1 FORWARD RCS MODULE, 2 AFT RCS SUBSYSTEMS IN PODS
38 MAIN THRUSTERS (14 FORWARD, 12 PER AFT POD)
THRUST LEVEL = 870 LB (VACUUM) EACH
I_{SP} = 280 SEC

6 VERNIER THRUSTERS (2 FORWARD & 4 AFT)
THRUST LEVEL = 24 LB EACH
I_{SP} = 260 SEC

PROPELLANTS:	OXIDIZER	FUEL
MAXIMUM LOADED	N_2O_4	MMH
AFT (2 PODS)	2976 LB	1860 LB
FORWARD	1488 LB	930 LB

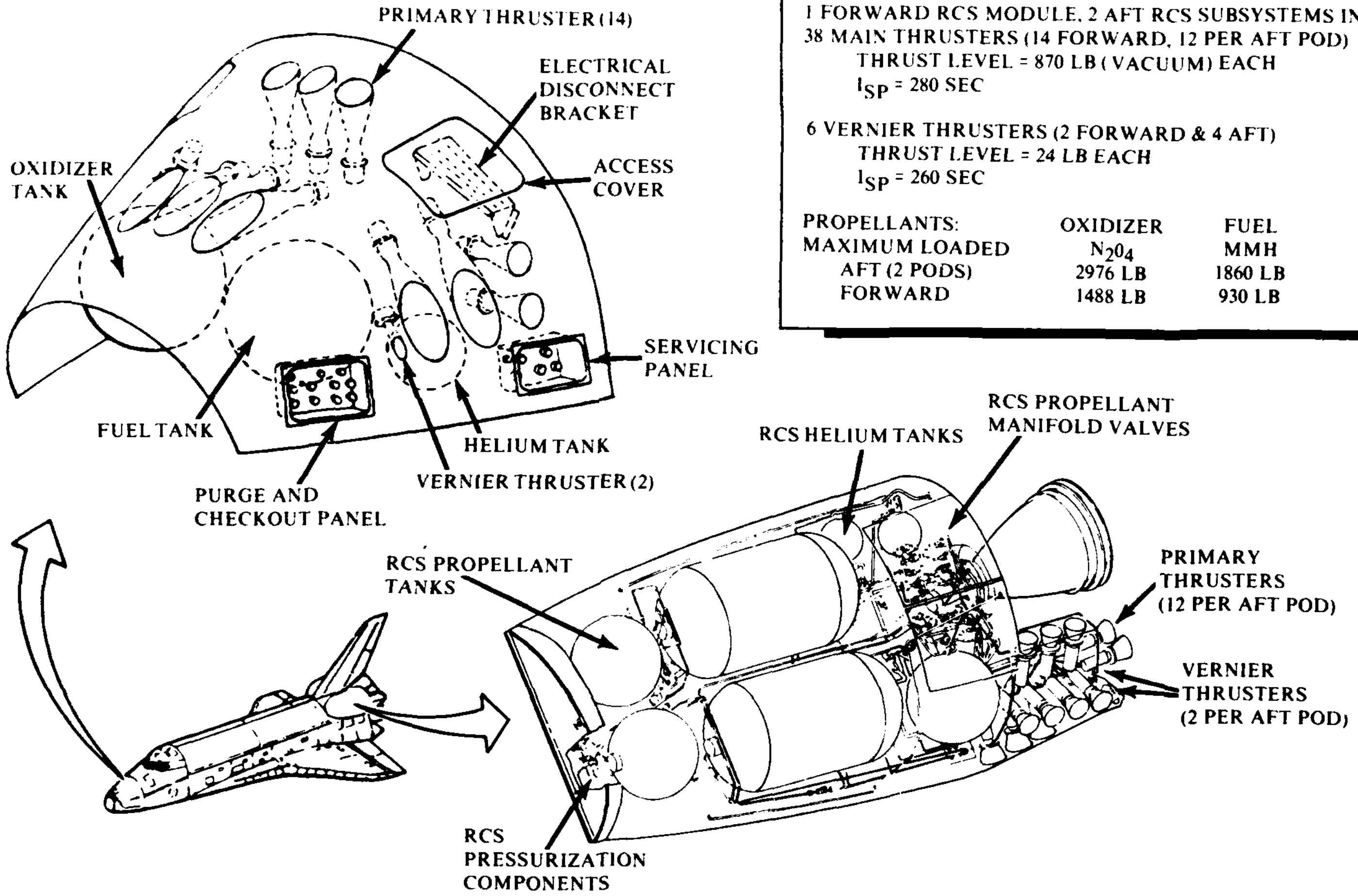


Figure 13-6. Orbiter Reaction Control Subsystem and thruster arrangement.

External Tank

The external fuel tank, the structural backbone of the system in its launch configuration, contains the propellants that feed the orbiter's three main engines at launch (see fig. 13-7). The tank is 50.6 meters (154.2 feet) long and 9 meters (27.6 feet) in diameter. It weighs 779,400 kilograms (1,718,265 pounds) loaded and 34,040 kilograms (75,045 pounds) empty. Actually three tanks in one, it consists of an oxygen tank located in the forward section, a rear hydrogen tank in the rear section, and an intertank that connects the two propellant tanks and houses instrumentation. A multilayered thermal coating covers the tank. It enables the tank to withstand extreme temperatures inside and outside the tank during prelaunch, launch, and early flight. Coating materials include a polyurethanelike foam on the external surface of the entire tank and an inner ablating compound on portions subject to high temperatures.

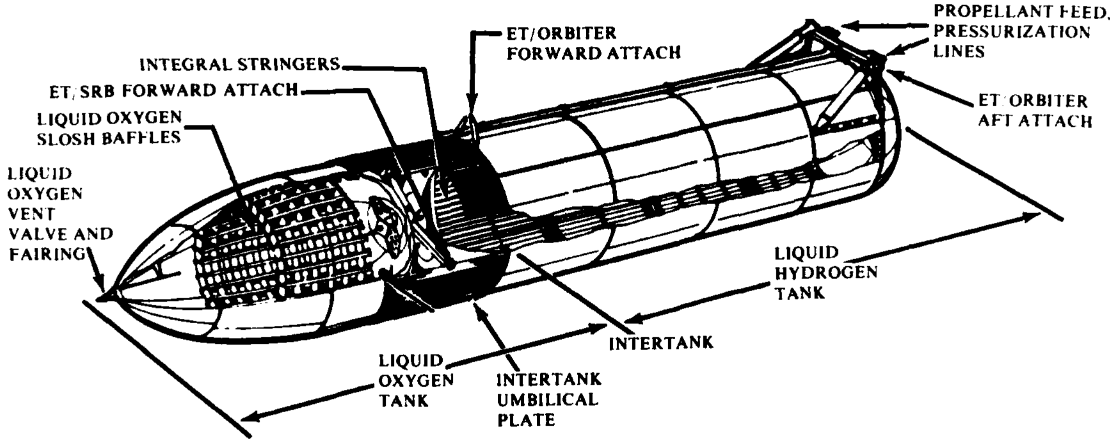


Figure 13-7. External fuel tank.

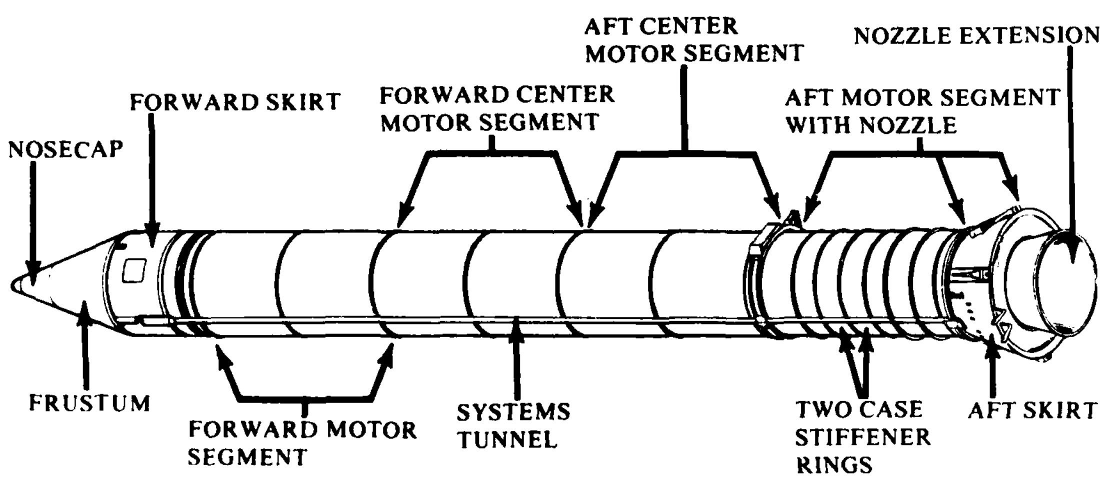


Figure 13-8. Solid rocket booster.

At launch, the propellants enter feedlines 43.2 centimeters (17 inches) in diameter and combine to fire the three main engines of the spacecraft. When the solid-propellant rocket motors separate at approximately 27 miles altitude, the orbiter, with its engines still burning, carries the external tank piggyback to a point 70 miles above the earth. As mentioned earlier, the external tank separates from the orbiter just before the craft achieves orbital velocity and falls in a preplanned trajectory into the Indian Ocean.

Solid-Propellant Rocket Boosters

The two solid-propellant rocket boosters used in each flight of the shuttle help to provide the initial thrust to boost the system with its payload from the launch pad to an altitude approximately 44 kilometers (27.4 miles). Each booster is 45.5 meters (149.1 feet) long and 3.7 meters (12.2 feet) in diameter (see fig. 13-8). It consists of six subsystems. These include the solid-propellant motor, vector control, separation, recovery, electrical, and instrumentation.

The heart of each booster is the motor, the largest solid-propellant rocket motor ever flown and the first designed for reuse. It consists of 11 segments joined together to make four loading segments. Propellant fills each segment. The segmented design permits easy fabrication, handling, and transportation. To help in steering the shuttle during flight, a hydraulically operated thrust vector control computer can move the exhaust nozzle in the rear segment of each motor a maximum of 6.65 degrees.

At burnout, pyrotechnic (explosive) devices separate the two boosters from the external tank, and four separation rockets in the forward nose frustrum of the shuttle and four on the rear skirt move the boosters away from the craft. A device for separating electrical connection with the external tank is part of the separation system. The booster recovery subsystem provides the means for controlling the velocity and altitude of the boosters during descent. Located in the forward section of each booster and within the nose cap, this subsystem consists of parachutes and location devices. These devices help in the ocean search and retrieval operations for each booster.

Following separation and entry into the lower atmosphere at approximately 4,700 meters (15,400 feet), a pilot, drogue, and three main parachutes slow each booster. When the booster lands in the ocean, air trapped in the booster causes it to float with the forward end projecting from the water. The booster has direction-finding beacons and lights to guide recovery craft. Then, ships tow the booster to shore and personnel ship motor segments to the manufacturer for refurbishment and reloading in preparation for another flight.

LAUNCH OPERATIONS AND GROUND SUPPORT

NASA will launch flights of the Space Shuttle from two locations, Kennedy Space Center in Florida and Vandenberg Air Force Base in California (see fig. 13-9 and 13-10). The government selected these sites to handle projected NASA and DOD payloads and to avoid flights over land masses during the launching of the shuttle. Currently NASA uses the Kennedy Space Center to launch payloads into orbits of 39 to 57 degrees inclination with respect to the equator. Eventually, NASA will use Vandenberg AFB for launches with orbital inclinations of 56 to 104 degrees. Range safety criteria restrict the launch directions (azimuths) from each launch site to prevent a released tank or booster from falling on a land area. The northernmost azimuth from the Kennedy Space Center is limited by the southeast portion of Newfoundland. The southernmost azimuth is limited by the Bahama Islands. Land constraints from Vandenberg Air Force Base include portions of Mexico to the southeast and the Hawaiian Islands to the southwest.

The direction of the earth's rotation has a significant bearing on capabilities for launching shuttle payloads. A due east launch (28.5 degrees inclination) using the earth's easterly rotation from Kennedy Space Center will permit a payload weighing up to 29,484 kilograms (65,000 pounds). A polar orbit launch from Vandenberg AFB, where the earth's rotation neither assists nor hinders the capabilities of the shuttle, will permit a payload up to 18,144 kilograms (40,000 pounds). The westernmost launch from Vandenberg AFB will allow transporting a payload up to

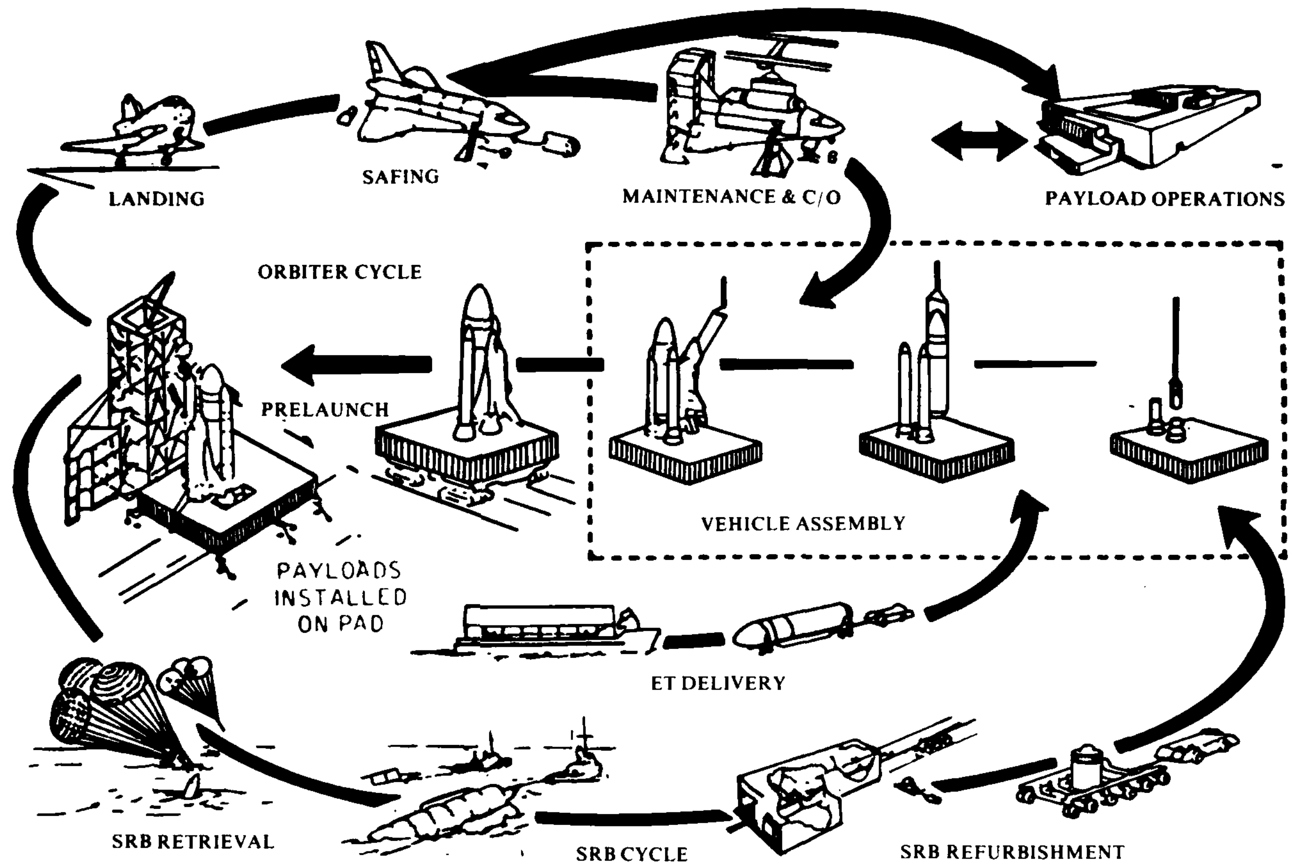


Figure 13-9. Kennedy Space Center Space Shuttle System ground flow.

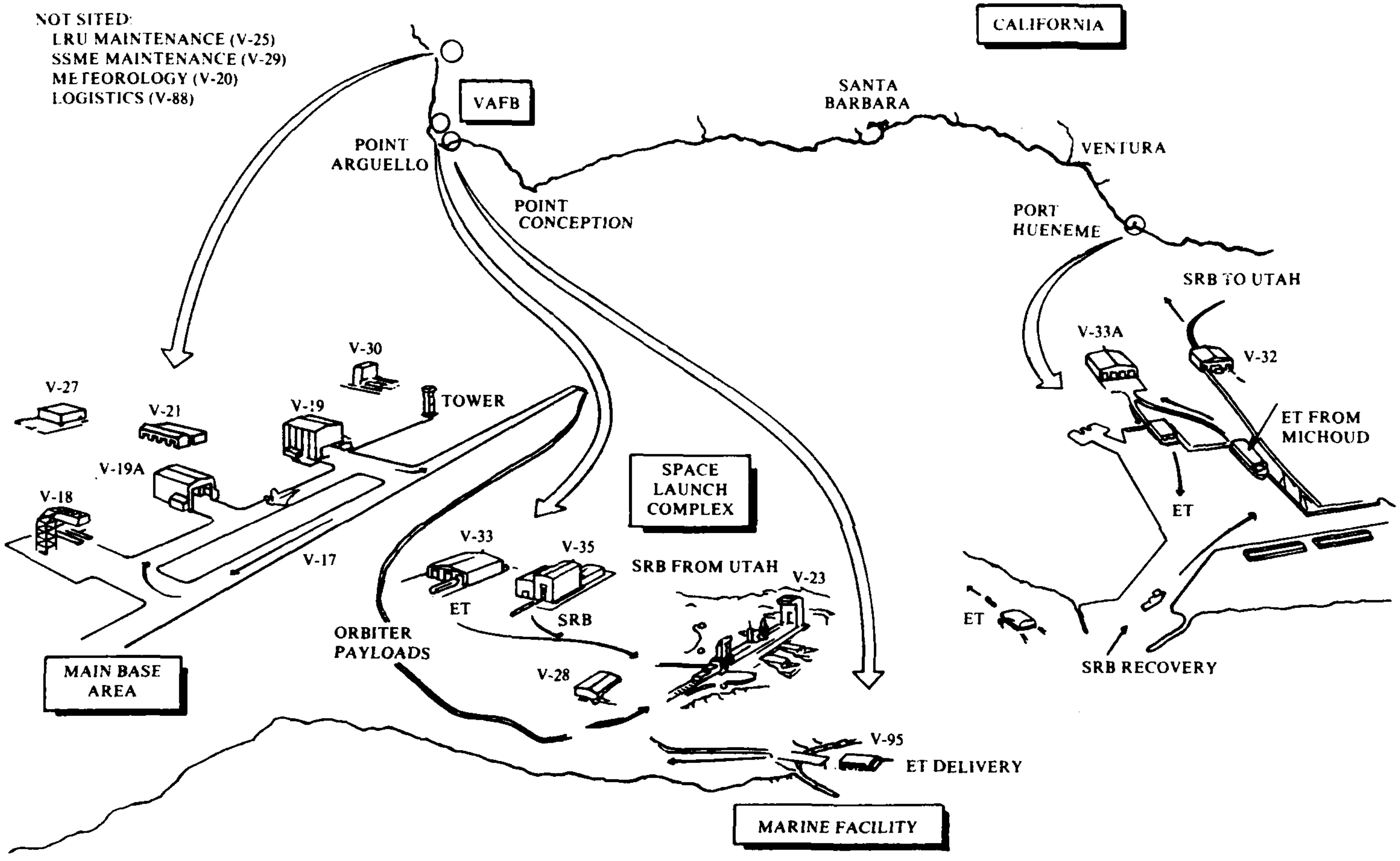


Figure 13-10. Vandenberg AFB ground support system.

14,515 kilograms (32,000 pounds) into orbit. This is possible because the earth runs counter to the westerly launch azimuth. (Figure 13-11 shows the launch azimuth and inclination limits from Vandenberg Air Force Base and Kennedy Space Center.)

When an orbiter returns to earth from its mission in space, it lands at the Kennedy Space Center (see fig. 13-9) located near the shuttle launch pad or at Vandenberg Air Force Base, California. The runway used by the shuttle is approximately twice the length and width of runways used at commercial airports. The shuttle runway is 4,572 meters (15,000 feet) long, 19.4 meters (300 feet) wide, and 40.5 centimeters (16 inches) thick at the center. The runway includes safety overruns of 305 meters (1,000 feet) at each end. The runway has a slope of 61 centimeters (24 inches) from the centerline to the edge. This slope of the concrete, plus small (0.63 centimeters or 0.25 inch) grooves cut into the runway every 2.85 centimeters (1.25 inches) across, provide rapid drainage and a more skid resistant surface. There is an aircraft apron, or ramp, 168 meters (550 feet) by 149 meters (490 feet) attached near the southeastern end of the runway. The mate/demate device located on the northeastern corner of the ramp lifts the orbiter for attachment to, or removal from, its Boeing 747 carrier aircraft for ferrying operations and provides movable platforms for access to certain orbiter components.

A sophisticated microwave scanning-beam landing system guides the orbiter automatically to a safe landing. Since the craft lacks propulsion during the landing phase, its high-speed glide must bring it in for a perfect landing the first time. An additional ground support system at Vandenberg Air Force Base, California (see fig. 13-10), will have facilities for assembly and checkout of payloads, the launch vehicle, and launch and recovery support activities. If possible, the support equipment, systems, and procedures will be compatible with the design of the ground systems at the Kennedy Space Center.

BIBLIOGRAPHY

Allaway, Howard. *The Space Shuttle At Work.* Washington D.C.: National Aeronautics and Space Administration, 1979.

Berry, F. Clifton. "Space is a Place." Interview with Lt Gen Richard C. Henry, Space Division, Air Force Systems Command. *Air Force.* (June 1982):36-42.

Boggs, Wallace H. and Samuel T. Beddingfield. "Moonport to Space." *Astronautics and Aeronautics.* (August 1982):28-41.

Rockwell International. *Space Shuttle System Summary.* Rockwell, Md.: Rockwell International Corporation, January 1978.

Schlitz, William P. "Supporting the Space Shuttle: The Blue-Suit Contribution." *Air Force.* (June 1982):44-51.

Space Shuttle 1980. Washington D.C.: United States Government Printing Office, January 1980.

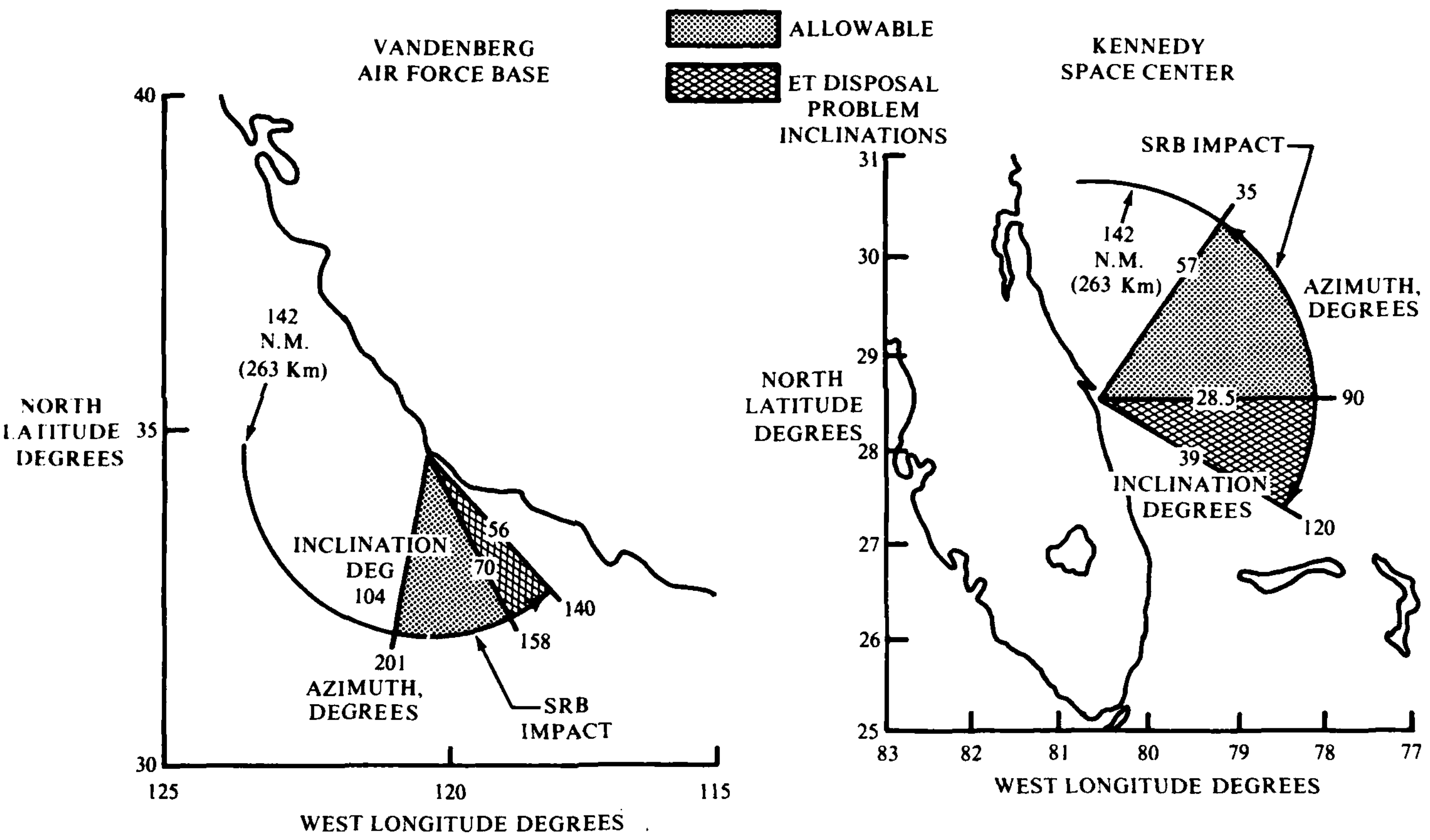


Figure 13-11. Launch azimuth and inclination limits from Vandenberg AFB and Kennedy Space Center.

Chapter 14

SPACE SUPPORT ORGANIZATIONS

There are numerous and diverse organizations throughout the United States that are active in space operations, research, and support. A complete listing would require a document as thick as this text, and would include such organizations as independent laboratories; colleges and universities; local, state, and federal government agencies; private companies; research foundations; and many more.

The intent of this chapter is to briefly outline the two major US organizations that provide the vast majority of our space activity, namely, the US Air Force and the National Aeronautics and Space Administration (NASA). Even these two organizations will not be described completely since each has components that are not involved in space or space-related activities and these organizations are in a constant state of flux.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

The National Aeronautics and Space Act of 1958 (72 Stat. 426; 42 U.S.C. 2451 et seq.) as amended established NASA. The principle statutory functions of NASA are to conduct research for the solution of problems of flight within and outside the earth's atmosphere; to develop, construct, test, and operate aeronautical and space vehicles; to conduct activities required for the exploration of space with manned and unmanned vehicles; to arrange for the most effective utilization of the scientific and engineering resources of the United States with other nations engaged in aeronautical and space activities for peaceful purposes; and to provide for the widest practicable and appropriate dissemination of information concerning NASA's activities and their results. Congress established these functions to carry out its policy of devoting space activities to peaceful purposes for the benefit of all mankind.

National Aeronautics and Space Administration Headquarters

Responsibility for planning, coordination, and control of NASA programs are vested in its administrator. Directors of NASA's field installations are responsible for execution of NASA's programs, largely through contract with research, development, and manufacturing enterprises. Government-employed scientists, engineers, and technicians conduct a broad range of research and development activities in NASA's field installations. These activities evaluate new concepts and phenomena and maintain the competence required to manage contracts with private enterprises.

Planning, direction, and management of NASA's research and development programs are the responsibility of the program offices that report to, and receive overall guidance and direction from, the administrator. The overall planning and direction of institutional operations at the field installation and management of agencywide institutional resources are the responsibility of the appropriate institutional associate administrator under the overall guidance and direction of the administrator.

The program offices include Aeronautics and Space Technology, Space and Science Applications, Space Transportation System, and Space Tracking and Data Systems. A brief description of the program responsibilities of these offices follows.

Aeronautics and Space Technology. The Office of Aeronautics and Space Technology is responsible for the conduct of programs to develop advanced technology to enable and enhance an aggressive pursuit of national objectives in aeronautics and space; to demonstrate the feasibility of this advanced technology, when necessary; to ensure its early use; to apply technology developed in the aeronautics and space programs; and to coordinate the application of NASA capabilities and facilities to programs of other agencies to accomplish energy-related research and development on a reimbursable basis. The office also coordinates the agency's total advanced research and technology program to ensure its overall adequacy and to avoid undesirable duplication.

Space Science and Applications. This office is responsible for NASA's efforts to understand the origin, evolution, and structure of the universe and the dynamic processes that shape the earth and its environment. It plans and implements programs leading to the study of earth as a planet and the solution of national problems. It coordinates the program between various government agencies, foreign interests, and the private sector in an effort to ensure the maximum exploitation of space science opportunities and the application of space science and technology to further knowledge and enhance life on earth.

In the achievement of its objectives, this office conducts research and development in astrophysics, earth and planetary exploration, environmental observations, life sciences, materials processing in space, communications, information systems, and Spacelab flight. The office utilizes the Space Transportation System automated spacecraft, sounding rockets, balloons, aircraft, and ground-based research in conducting the program. It is responsible for NASA contacts with the Space Science Board of the National Academy of Sciences and other science advisory groups.

Space Transportation Systems. This office is responsible for advancing the Space Shuttle and carrying out space transportation and other associated programs. It is responsible for planning, directing, and acquiring all elements of the space transportation system. The office manages, directs, and coordinates all US civil launch capabilities and US Spacelab development, procurement, and operations. It has the responsibility for developing and implementing necessary policy for all users of the space transportation system; and developing detailed plans, schedules, and resource requirements for space flight program objectives.

Space Tracking and Data Systems. This office is responsible for providing tracking, data acquisition and processing, and communication support to all NASA flight projects. Types of flight projects supported include sounding rockets, research aircraft, earth orbital and suborbital missions, lunar and planetary spacecraft, deep space probes, and the Space Shuttle program. To provide such support the office plans, implements, and operates a worldwide network of fixed and mobile tracking facilities for acquiring data, a centralized control center complex for directing missions, a large-scale data processing facility, and an integrated communications network for data transmission.

Field Installations

A brief description of the principal and supporting roles of NASA's field installations follows.

Ames Research Center, Moffett Field, California. A list of this center's areas of responsibility include fundamental aerodynamics; shorthaul aircraft technology; rotocraft technology; computational techniques; human factors research; computer science research; aeronautical flight research; remotely-piloted vehicle research; flight simulation; technical support to military aviation; airborne science and applications; planetary entry; planetary spacecraft development and mission operations; life sciences; space transportation passenger selection criteria; astronomical observation techniques; shuttle orbiter development support; and advanced space vehicle configurations technology.

Goddard Space Flight Center, Greenbelt, Maryland. This center is responsible for earth orbital spacecraft development and flight operations; tracking and data acquisition systems and support operations; spacelab payloads; space physics and astronomy payloads and science; and upper atmospheric research. Its responsibility includes applications and research development in weather and climate, earth dynamics, earth resources, communications; information systems technology; sounding rocket and payload development, procurement, and operations; balloons; launch vehicle procurement; planetary science; and sensors and experiments in environmental monitoring and ocean dynamics.

Jet Propulsion Laboratory, Pasadena, California. This laboratory is operated under contract by the California Institute of Technology. Its research areas include planetary spacecraft development and mission operations; tracking and data acquisition for deep space operations; lunar/planetary science; upper atmospheric research; science and applications spacecraft development; earth and ocean dynamics; teleoperator technology; guidance and control technology; energy technology for civil applications; space physics payloads and science; space astronomy payloads and science; sensor and data acquisition systems and technology; and information systems technology.

Lyndon B. Johnson Space Center, Houston, Texas. This center is responsible for manned space vehicles and supporting technology; flight operations including flight control; life sciences; lunar and planetary geosciences; earth resources surveys; technology experiments in space; and energy systems.

John F. Kennedy Space Center, Kennedy Space Center, Florida. This center works with expendable vehicle launch operations; Space Transportation System (STS) ground operations; and STS sustaining engineering. It must collaborate with such elements of the Department of Defense as the Eastern Test Range and Corps of Engineers to avoid unnecessary duplication of launch facilities, services, and capabilities.

Langley Research Center, Hampton, Virginia. This center carries out research on long-haul aircraft technology; general aviation aircraft/commuter aircraft technology; fundamental aerodynamics; acoustics and noise reduction; aerospace vehicle structures and materials; aerothermodynamics; sensors and instruments; flight crucial electronics and aircraft control systems; technical support to military aviation; advance space vehicle configurations technology; sensor and data acquisition technology; technology experiments in space; atmospheric services technology; hypersonic propulsion systems; planetary entry technology; computational fluid dynamics; upper atmospheric research; and launch vehicle procurement.

Lewis Research Center, Cleveland, Ohio. This center's responsibilities include research on air breathing propulsion systems; engine materials and structures; structural dynamics; turbomachinery aerodynamics combustion; propulsion controls; power transmission; military support; tribology; internal engine computational fluid dynamics; and high-temperature engine instrumentation. Other areas of responsibility are launch vehicle procurement; space propulsion systems technology; space energy processes and systems technology; energy technology and applications; high-power communications; and ground transportation system research.

George C. Marshall Space Flight Center, Huntsville, Alabama. This center works with propulsion systems; manned space vehicle development; Spacelab mission management and payload definition; specialized automated spacecraft; space processing; space vehicle structures and materials; energy technology and applications; and satellite power system definition.

National Space Technology Laboratories, Bay St. Louis, Mississippi. These laboratories conduct research on static test firing of large space and launch vehicle-engines. They house certain

environmental research and earth resources activities of NASA and other governmental agencies as well, with emphasis on the use of space technology and associated managerial and technical disciplines.

UNITED STATES AIR FORCE

The National Aeronautics and Space Act of 1958 begins with the statement:

> The Congress hereby declares that it is the policy of the United States that activities in space should be devoted to peaceful purposes for the benefit of mankind.
>
> The Congress declared that the general welfare and security of the United States require that adequate provision be made for aeronautical and space activities. The Congress further declares that such activities shall be the responsibility of, and shall be directed by, a civilian agency exercising control over aeronautical and space activities sponsored by the United States, except that activities peculiar to or primarily associated with the development of weapon systems, military operations, or defense to the United States (including the research and development necessary to make effective provision for the defense of the United States) shall be the responsibility of, and shall be directed by, the Department of Defense; and that determination as to which such agency has responsibility for and direction of any such activity shall be made by the President

The US Air Force has conducted, and is conducting, most of the Department of Defense (DOD) space activities. Although DOD space missions are not exclusively the responsibility of the US Air Force (the US Navy's Fleet Satellite Communications handles some), the US Air Force is DOD's exclusive agent for the launching of all DOD payloads and for the STS (Space Shuttle)

The Space Command

The Space Command was established as a major command on 1 September 1982. It is an operational command designed to carry out national space policy that enhances military capabilities for defense of the United States.

Mission and tasks. The Space Command has a current operational mission that entails the mission of aerospace defense and a far-reaching agenda for future space activities. It will develop space strategies and operational concepts that address such factors as orbit selection, position management, orbital force structure, and others. It will also promote a comprehensive documentation of the Soviet space threat.

Space Command will handle the possible future task of operating the US antisatellite system when it is deployed. This system will represent a necessary element in deterrence. In the Cheyenne Mountain Complex near Colorado Springs, there is a strong nucleus of space expertise as well as facilities, communication, and computer capabilities to carry out space missions. This nucleus, combined with resources acquired from the Air Force Systems Command, Strategic Air Command, and Air Force Communications Command, will provide the initial ingredients of the Space Command.

The operational requirements of commanders in the field will be a major concern of the Space Command. The command will incorporate space assets into joint-service exercises. Present operational plans call for land, air, and sea contingencies. The Space Command will integrate space activities into these plans.

The Space Command will actively influence Air Force technology development efforts by acting as a catalyst between the operational and technological worlds. In this role, the command will review existing programs and add impetus to the development of promising new technologies. A current technology review effort involves high-energy lasers and associated laser development programs.

The survivability of space systems will be a major area of the Space Command effort. Military forces are dependent on space systems. These must be available when needed in crises and conflicts. The Space Command's survivability efforts will be directed toward vulnerable nodes such as launch sites; tracking, telemetry, and commanding facilities; associated comm ications

network; and spacecraft.

Space medicine is another important challenge. Space sickness and the malfunction of the spacesuits during the fifth shuttle flight underscore the effect that problems with the space environment and life support systems have in our ability to operate in space. The Space Command surgeon will press for research efforts to understand the space environment and for design improvements to increase the use of people in space.

The availability of people with the requisite technical training and education is another concern. In the future, one of the biggest limiting factors could be the lack of skilled people in the space disciplines. As a priority task, the Space Command will oversee training for Air Force space activities and monitor and create added opportunities for career development to ensure qualified people in sufficient numbers are available to meet the Air Force's needs. To meet this need, a space systems major at the Air Force Academy has been implemented. The Academy implemented it for school year 1984. The Air Force Institute of Technology (AFIT) currently offers a master's degree in space operations management at Wright-Patterson Air Force Base, Ohio, and the Space Command will be promoting enrollment to meet future needs.

It is clear that space is the place for the future, and that more military missions can be supported from space. Space-based systems could alter the basic premises of national strategy. In the nuclear era, it has been all offense—land-based bombers and missiles and missile-carrying submarines. Now, space may provide viable options—alternate investment opportunities—to change this equation and put greater emphasis on strategic defense.

The establishment of Space Command is occurring at the right time, in the right place, with the right people, and the entire Air Force is excited about it. The Space Command is a major step toward meeting the president's policy goal of having a space program that will strengthen national security. The Space Command motto, Guardians of the High Frontier, symbolizes the importance of space to the national security, and underscores our commitment to leadership in this expanding medium.

Organization. In addition to the space facilities in Cheyenne Mountain the Space Command will manage and operate the worldwide space surveillance and missile warning radars and optical systems. The Air Force has assigned the Satellite Early Warning system and the Defense Meteorological Satellite program to the Space Command as well.

The command will be responsible for the Consolidated Space Operations Center (CSOC) near Colorado Springs when completed. The CSOC will control operational spacecraft and plan, manage, and control all DOD shuttle flights. It will be the focal point of military space operations.

The recently activated Space Technology Center in Albuquerque, New Mexico, is responsible for unique space technical disciplines. It will ensure technological leadership in areas critical to US security. Now the Air Force can process all space systems from basic technology through development, acquisition, launch, and on-orbit checkout by the Space Division of Air Force Systems Command to on-orbit control by Space Command.

Space Command gives the Air Force an operational command to manage, control, and protect space assets. It will provide a much closer relationship between the research and development community and the operational world. Space Command will provide a focus for centralized plans, consolidated requirements, and the needs of operational commanders. It will provide the operational pull to go along with the technology push that has been dominant in the space world.

North American Aerospace Defense Command

The United States and Canada continue a lasting partnership in the defense of North America. This partnership has evolved into a deep and abiding friendship providing the essential ingredient to the collective security of our two countries. The organization that implements this defense arrangement is the North American Aerospace Defense Command (NORAD).

NORAD is the first binational command to be established within North America. It provides

the cooperative defense planning and organizing wherein the governments of Canada and the United States place military forces under a single commander—CINCNORAD. The NORAD organization melds multiservice and multinational elements into an effective combat force.

CINCNORAD reports to the president through the Joint Chiefs of Staff and secretary of defense and to the Canadian prime minister through the Canadian chief of defense staff and minister of national defense. The commander's duties and responsibilities are parallel in the United States and Canada. The NORAD headquarters has the normal six-deputate joint staff. Canadian Forces and members of the United States Air Force, Army, Navy, and Marine Corps occupy key NORAD staff positions. The NORAD commander heads a binational joint force comprised of integrated components of a US specified command and the air defense elements of the Canadian Forces Air Command. In addition, the NORAD commander is also commander in chief, Aerospace Defense Command (ADCOM) the US specified command and commander of Space Command. The ADCOM commander's responsibilities include operational command of the aerospace defense of the CONUS and Alaska under circumstances requiring unilateral aerospace defense actions by the United States. In brief, ADCOM is NORAD less the Canadians.

Air Force Systems Command

Air Force Systems Command, established 1 April 1961, is responsible for the rapid advancement of aerospace technology and its adaptation into operational aerospace systems. The command is organized into divisions, centers, national test ranges, laboratories, and research and development offices.

Product divisions develop, test, and procure systems and equipment. They contain system program and project offices and operate as subcommands. These may include a system engineering and technical integration contractor, which operates as a nonprofit corporation established to advise the Air Force on system engineering and technical direction. The nonproduct divisions analyze and evaluate technological threats; conduct biotechnology research, development, and education programs; and provide contracting liaison.

The centers and two national ranges provide development, test, and evaluation facilities for the command. They have other specialized facilities such as rocket test stands, wind tunnels and simulators, sled test tracks, and electronic test ranges. The two national ranges have the capability to form a single global tracking network for intercontinental ballistic missiles (ICBMs), space satellites, launch vehicles, and space probes.

The laboratories and Air Force Office of Scientific Research (AFOSR), Bolling Air Force Base, D.C., in addition to providing research and development support to system and equipment acquisition programs of the product divisions, constantly seek to improve the foundation of advanced technology. Special funds exist to support opportunistic research and development that, because of timing or other circumstances, is not funded during the normal budgeting cycle. The AFOSR grant authority is unique in the Air Force.

Space Division. The Space Division, headquartered in Los Angeles, California, with launch and tracking systems worldwide, has the responsibility for the space systems. This responsibility encompasses engineering, test, program management, installation, launch, on-orbit tracking, and command and control.

The Space Division manages the design, launch, and on-orbit command and control of military spacecraft. The unit plays a major development role in the ground terminal segment of space systems. Today's space systems support communications, early warning, meteorology, surveillance, and navigation needs of operational commands. The Space Division overseas the spacecraft technology development programs and space defense programs as well.

Space launch boosters are a major element of the mission. In addition to being responsible for current expendable boosters such as Atlas and Titan, the Space Division is the DOD's manager for military activities in the national space transportation system. While the National Aeronautics

and Space Administration (NASA) manages the overall space shuttle program, the Space Division is responsible for integration of military payloads and the operation of the west coast spaceport at Vandenberg Air Force Base, California. It is responsible for developing the inertial upper stage to propel payloads from low-earth orbits of the shuttle to high-altitude orbits as well.

Space Division units include the Space Missile Test Organization, the Air Force Satellite Control Facility, and the Manned Space Flight Support Group.

Space and Missile Test Organization. The Space and Missile Test Organization (SAMTO), through its two operational units, the Eastern Space and Missile Center and the Western Space and Missile Center, provides flight-test management for DOD-directed ballistic missile and space programs. It manages more than 20 satellite launches a year from Vandenberg AFB, California, and Cape Canaveral, Florida, as well as a variety of ICBM ballistic tests. The centers operate the Eastern and Western Test Ranges, which stretch around the world. SAMTO supervises a vast array of data-gathering sites scattered throughout its ranges in support of its test programs and those of the Strategic Air Command, National Aeronautics and Space Administration, and other government agencies.

Air Force Space Technology Center. The Air Force Space Technology Center (AFSTC) is located at Kirtland Air Force Base, New Mexico, and operates under the Space Division. AFSTC was activated in October 1982 as the Air Force focal point for space technology planning and development. It coordinates technology programs for space missions, ensuring correct technological choices and investment strategy.

The center manages three AFLC laboratories. These are the Air Force Weapons Laboratory, Kirtland Air Force Base, New Mexico; Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts; and Air Force Rocket Propulsion Laboratory, Edwards Air Force Base, California. The laboratories perform research and development for the Air Force and nonmilitary organizations in their respective disciplines. The center integrates the space technology efforts of the laboratories to explore military space capabilities and the needs of future space systems. The center is a focal point for information about the scope and progress of space-related developments in electronics hardening and laser research, rocket propulsion, and the earth and space environments.

The center, in conjunction with other Space Division elements and Space Command, correlates research results with systems demonstration needs and identifies key technology areas for long-range plans. The Space Technology Center cooperates with the National Aeronautics and Space Administration and other military agencies on joint efforts to develop routine, reliable, and survivable spacecraft.

Air Force Satellite Control Facility. The Air Force Satellite Control Facility (AFSCF) provides orbital support for the research, development, test and evaluation, and operation of DOD space systems and, as required, NASA and NATO programs. From its headquarters at Sunnyvale, California, it operates a worldwide network of tracking stations.

Manned Space Flight Support Group. This unit is located at the Johnson Space Center, Houston, Texas. It is responsible for developing the capability to plan and control DOD space transportation system missions and to ensure that those missions are secure. It will manage the acquisition phase of the Shuttle Operations and Planning Center portion of the Consolidated Space Operations Center. The group will train personnel to directly support the command and control of DOD space shuttle missions and transition those personnel to the Shuttle Operations and Planning Center.

Electronic Systems Division. This division, located at L. G. Hanscom Field, Massachusetts, is responsible for developing, acquiring, and delivering electronic systems and equipment for the

command, control, and communications functions of aerospace forces. These systems take many forms, such as undersea communications cable around the Indochina peninsula, line-of-sight and tropospheric scatter communications throughout the Mediterranean, the underground NORAD combat operations center, long-range radars to warn of missile and aircraft attack, the air defense control net for the North American continent, equipment for improved weather forecasting, the free world's satellite detection and tracking network, and experimental satellites for the development of communications satellite technology.

Rome Air Development Center. The Rome Air Development Center (RADC), located at Griffiss AFB, New York, is the principal organization charged with Air Force research and development programs in information sciences; intelligence; and command, control, and communications. RADC mission areas include communications; electromagnetic guidance and control; surveillance of ground and aerospace objects; intelligence data collection and handling; information systems technology; ionospheric propagation; solid-state sciences; microwave physics; and electronic reliability, maintainability, and compatibility.

Arnold Engineering Development Center. Located at Arnold AFS, Tennessee, the Arnold Engineering Development Center (AEDC), currently valued at more than $940 million, contains the most advanced and largest complex of aerospace flight simulation test facilities in the free world. Its mission is to ensure that aerospace hardware, which includes aircraft, missiles, spacecraft, jet and rocket propulsion systems, and other components, will work right the first time they fly. The three major facilities of the center contain some 40 test units in which the engines can simulate flight conditions from sea level to altitudes of 1,000 miles and from subsonic velocities to more than 20,000 miles per hour. Equipment being tested ranges in size from small-scale models to full-scale vehicles with propulsion systems installed and operating.

Air Force Wright Aeronautical Laboratories. The Air Force established the Wright Aeronautical Laboratories (AFWAL) at Wright-Patterson AFB, Ohio, on 1 July 1975. It consists of a command and staff element with four laboratories: Air Force Avionics Laboratory, Air Force Aero Propulsion Laboratory, Air Force Flight Dynamics Laboratory, and Air Force Materials Laboratory.

The mission of the AFWAL is to plan and to execute USAF exploratory development, advanced development, and selected research and engineering development programs for flight vehicles, aeropropulsion, avionics, and materials, and the US Air Force manufacturing methods program. It provides support within its areas of technical competence for planning, developing, and operating aerospace systems to Air Force, Department of Defense (DOD), and other government agencies.

BIBLIOGRAPHY

AU-23, US Air Force Commands and Agencies: Basic Information, 1983-84. Maxwell AFB, Ala.: Air University Press, 1983.
United States Government Manual, 1982/83. Washington, D.C.: US Government Printing Office, 1 July 1982.

Chapter 15

SPACE POLICY AND DOCTRINE

Space policy defines in broad terms the basic goals and principles of the US space program. Space policy is shaped by national interests and security objectives and constrained by fiscal considerations and US objectives under international law. Perhaps policy formulation is the most critical element of the national planning process because it provides the framework for the subsequent development of military space strategy and the identification of future system requirements.

This chapter examines the basic tenets of US space policy, the national interests and security objectives that shape it, and the international legal regime that constrains it. The chapter summarizes DOD and Air Force space policies, which are derived from national space policy, and concludes with an analysis of those fundamental doctrinal principles that guide the conduct of military space activities.

NATIONAL INTERESTS

The nation's fundamental national interests have remained constant while the threats and circumstances facing the United States, and the capabilities needed to meet them, have changed over time. Though variously defined, they are four in number.

The first of these is national survival. It is the most basic of the national interests. Without it, no other interest has relevance. This includes preservation of freedom, political identity, and the institutions that provide their foundation (that is, the Constitution and the rule of law). The second is territorial integrity. This includes the protection of the territory of the United States, its citizens, and its vital interests abroad from armed attack. The third fundamental interest is economic well-being. We must be assured of access to foreign markets and overseas resources to maintain the strength of the US industrial, agricultural, and technological base. The fourth is a favorable world order. This means an international order that is not inimical to US interests or democratic institutions, economic development, and self-determination. Like the land, sea, and air mediums, space conceptually gives a new opportunity to further and protect US national interests in all four areas.

SPACE LAW

Though shaped by national interests, the development of space policy is constrained, nevertheless, by the US objectives under international law. Typically, but not exclusively, these obligations are embodied in bilateral or multilateral treaties signed and ratified by the United States. As such they are binding and as much a part of US law as provisions of the Constitution or statutes enacted by Congress.

Table 15-1 summarizes those international agreements that presently constrain US military activities in space and describes the nature of each constraint. Briefly stated, the following activities are prohibited or otherwise constrained as indicated:

- The appropriation by claim of sovereignty, use or occupation, or any other means, of any portion of outer space to include the moon and celestial bodies.

Table 15-1.
International Agreements that Limit Activities in Space

Agreement	*Prohibition/Constraint*
United Nations Charter (1947)	Does not specifically mention space; however, its provisions are made applicable to outer space by the Outer Space Treaty. Article 2(4) of the Charter prohibits states from threatening to use, or actually using, force against the territorial integrity of political independenece of another state. This prohibition is qualified, however, by Article 51 which recognizes a state's inherent right of individual or collective self-defense in the event of armed attack. Customary international law has expanded this qualification to include a state's right to defend its territorial integrity and political independence against impermissible coercion—a right that in turn has been interpreted by many states, the United States and Soviet Union included, to legitimize "anticipatory" self-defense, or the right to act in self-defense to remove a danger or threat of imminent armed attack. It is this right of "anticipatory" self-defense that the United States and Soviet Union have used to legitimize the development of antisatellite systems.
Limited Test Ban Treaty (1963)	Treaty banning nuclear weapon tests in the atmosphere, in outer space and under-water. States may not conduct nuclear weapon tests or other nuclear explosions (that is, peaceful nuclear explosions) in outer space or assist or encourage others to conduct such tests or explosions. (Article I)
Outer Space Treaty (1967)	Treaty on principles governing the activities of states in the exploration and use of outer space, including the moon and other celestial bodies. States may not place in earth orbit, install on celestial bodies, or station in space in any other manner weapons of mass destruction (generally defined as nuclear, chemical, and biological). (Article IV) States may not build military bases, installations, and fortifications; test weapons of any kind; or conduct military maneuvers on the moon or other celestial bodies. Celestial bodies are not completely demilitarized, however, because the treaty does not permit the "use of military personnel for scientific research or for any other peaceful purpose" as well as the "use of any equipment or facility necessary for peaceful exploration," which presumably includes military equipment and facilities. (Article IV) If a state plans an activity which could cause potentially harmful interference with the activities of other states, the state is obligated to consult with those states possibly affected prior to initiating the activity. Conversely, if a state believes that an activity by another state may cause potentially harmful interference, it may request consultations. However, there is no requirement that the state planning the activity must agree to consultations. (Article IX) States must carry out their exploration and use of space in such a way as to avoid harmful contamination of the moon or other celestial bodies as well as to avoid the introduction of extraterrestrial matters that could adversely affect the environment of the earth. (Article IX) Astronauts of one state are required to render all possible assistance to astronauts of other states either in outer space or on celestial bodies. (Article V) This requirement presumably refers only to accidents, situations of distress, or other emergencies, and presumably would require the assisting astronauts to provide shelter to the affected astronauts. States are asked to inform the secretary general, the public, and the international scientific community of the nature, locations, and results of any activities they conduct in outer space. This is not an absolute requirement; rather, it is caveated "to the greatest extent feasible and practicable." (Article XI) As might be expected, the details of military activities in space are typically obscured by launching states.

Table 15-1 (Continued)
International Agreements that Limit Military Activities in Space

Agreement	*Prohibition/Constraint*
Outer Space Treaty (Cont'd)	A state's space stations, installations, equipment, and vehicles located on the moon may be visited by representatives of other states provided that reasonable advance notice is given and that the visiting state permits reciprocal visits. The purpose of advance notices is to permit consultations over the appropriate timing of the visit so as to ensure the safety of the occupants and that the visit will not interfere with the normal operation of the facility. (Article XII)
Antiballistic Missile (ABM) Treaty and agreed statements thereto (1972)	Treaty between the United States of America and the Union of Soviet Socialist Republics on the limitation of antiballistic missile systems. The United States and the Soviet Union may not develop, test, or deploy space-based ABM systems or components. (Article V) This is a comprehensive ban. It includes currently understood ABM technologies (that is, interceptor missiles) as well as those concepts based on technologies yet to be fully developed or understood (that is, directed energy weapons). The United States and the Soviet Union may not interfere with the national technical means of verification of the other providing that such systems are operating in accordance with generally recognized principles of international law and are used to assure compliance with treaty provisions. (Article XII) A similar provision is contained in the Threshold Test Ban Treaty (Article II) and the SALT II agreement (Article XV). A case presumably could be made that either country was free to interfere with a "national technical means of verification" system that was operating in violation of generally recognized norms of international law <u>or</u> with a reconnaissance system that was not being used to verify treaty compliance—admittedly a difficult distinction to make—whether or not such a system was operating in accordance with the norms of international law. If ground-based ABM systems based on other physical principles and including components capable of substituting for conventional ABM interceptor missiles, launchers, or radars are created in the future, the parties agree to discuss limitations on such systems. (agreed statement D)
Convention on Registration (1974)	Registration of objects launched into outer space. States must notify the UN Secretary General of all objects launched into space and must provide the following information: name of the launching state; date of launch; territory from which the object was launched; a designator or registration number for the object; basic orbital parameters to include nodal period, inclination, apogee, and perigee; and the general function of the object. (Articles II, III, and IV) States must also notify the secretary general when the object is no longer in space. (Article IV)
Environmental Modification Convention (1980)	Convention on the prohibition of military or any other hostile use of environmental modification techniques. States may not use environmental modification techniques to destroy, damage, or injure another state if such usage has widespread (on the order of several hundred square kilometers), long-lasting (a season or more), and severe (serious or significant disruption or harm to human life, natural and economic resources or other assets) effects. Environmental modification techniques are defined as any technique for changing the dynamics, composition, or structure of the earth or outer space through the deliberate manipulation of natural processes. (Article I)

- Threatening to use or actually force against the territorial integrity and political independence of another state.
- Testing nuclear weapons or other nuclear explosive devices, even peaceful nuclear devices.
- Placing in earth orbit, installing on celestial bodies, or stationing in space in any other manner weapons of mass destruction (generally defined as nuclear, chemical, or biological).
- Building military bases, installations, or fortifications on the moon or other celestial bodies.
- Testing weapons of any kind on the moon or other celestial bodies.
- Conducting military maneuvers on the moon or other celestial bodies.
- Developing, testing, or deploying space-based antiballistic missile (ABM) systems or components (for example, missions, launchers, or radars).
- Developing future ground-based ABM systems based on physical principles other than and including components capable of substituting for conventional ABM interceptor missiles, launches, or radars without first discussing with the Soviet Union possible limitations on such systems.
- Interfering with Soviet national technical means of verification provided such systems are operating in accordance with generally recognized principles of international law and are in fact being used to verify provisions of the ABM Treaty, SALT, and Threshold Test Ban Treaty.
- Initiating activities that could cause harmful interference with the activities of other states without first consulting with those states.
- Causing harmful contamination of the moon or other celestial bodies.
- Using environmental modification techniques to destroy, damage, or injure another state.
- Launching space objects without notifying the Secretary General of the United Nations.

One of the fundamental truths of international law is that if an act is not prohibited specifically, then that act is permitted. Therefore, though the list of prohibited acts is sizable, in the aggregate there are very few legal restrictions on the use of space for nonaggressive military purposes. Therefore, international law permits implicitly the performance of traditional military support functions such as surveillance, reconnaissance, navigation, meteorology, and communications. It permits the deployment of military space stations; the testing and deployment in earth orbit of nonnuclear, non-ABM weapon systems, the use of space for individual and collective self-defense and any conceivable activity not specifically prohibited or otherwise constrained.

A second and equally fundamental truth is that, in most instances, treaties are designed to regulate activities between the signatories during peacetime only. Unless the international agreement states clearly that its provisions are designed to apply or become operative during hostilities, or the signatories can deduce this from the treaty's content, they must presume that armed conflict will result in the suspension or termination of the applicability of a treaty's provisions. This is particularly true for treaties whose purpose is to disarm or limit quantities of arms maintained by signatories. Thus, in time of hostilities, the scope of permissible military space activities broadens significantly.

NATIONAL SPACE POLICY

The origin and development of the nation's space policy is interesting and extremely important. If we are to understand the evolution of the present policy and predict the future course of space policy, we must begin with a brief survey of its origin and evolution.

Early Policy

The launch of Sputnik I on 4 October 1957 had an immediate and dramatic impact on the formulation of US space policy. Though the military had expressed an interest in space technology as early as the mid-1940s, a viable US satellite program failed to emerge for several reasons. These include intense interservice rivalry; military preoccupation with the development of ballistic missiles that prevented a sufficiently high funding priority from being assigned to proposed space systems; and, perhaps most importantly, national leadership that failed initially to appreciate the strategic and international implications of emerging satellite technology, and when it did, was committed to an open and purely scientific space program.

Sputnik I changed all that. It resulted in the first official US government statement that space indeed was of military significance. This was a 26 March 1958 statement by President Eisenhower's science advisory committee that the development of space technology, in addition to being required by human curiosity, scientific knowledge, and the maintenance of national prestige, was important for the "defense of the United States" as well. Congress recognized quickly that space activity was potentially vital to the national security. The following language of House Resolution 1770 reflected this.

> This country is not unmindful of what these Soviet achievements (in space) mean in terms of military defense.... Ballistic missiles already travel for a considerable part of their path through near outer space and can arrive virtually without warning to deliver their devastating thermonuclear warheads. The United States must have a strong capability in the use of outer space, both as a deterrent to military use of space vehicles against this country and as an aid in developing antimissile techniques. Satellite (operations) will have important implications for guarding the peace. On the one hand, they are adjuncts to weapon systems related to the deterrent power, and, on the other, they represent important techniques for inspection and policing, in accordance with any disarmament scheme which may be negotiated in the years to come.*

The first official national space policy was embodied in the National Aeronautics and Space Act of 1958. It declared that the policy of the United States was to devote space activities to peaceful purposes for the benefit of all mankind, mandated separate civilian and national security space programs, and created a new agency, the National Aeronautics and Space Administration (NASA), to provide direction to and exercise control over all US space activities except those "peculiar to or primarily associated with the development of weapons systems, military operations, or the defense of the United States." The Department of Defense (DOD) was to be responsible for, and to direct, these latter activities. A legislative basis for DOD responsibilities in space was provided early in the space age. The act established a mechanism for coordinating and integrating military and civilian research and development, encouraged significant international cooperation in space, and called for preserving the role of the United States as a leader in space technology and its application.

The policy framework for a viable space program was thus in place. In fact, the principles enunciated by the National Aeronautics and Space Act, which included peaceful focus on the use of space, separation of civilian and national security space programs, emphasis on international cooperation, and preservation of a space role, have become basic tenets of the US space program. All presidential space directives issued since 1958 have reaffirmed these basic tenets. However, what was missing was a space program of substance. The Eisenhower Administration's approach to implementing the new space policy can be characterized as conservative, cautious, and constrained. Early DOD and NASA plans for advanced manned space flight programs were disapproved consistently. Instead the administration preferred to concentrate on unmanned, largely scientific missions and to proceed with those missions at a measured pace. It was left to subsequent administrations to give the policy substance.

*1958 US Code Congress and Administration News, 3161-3162

Intervening Years

Two presidential announcements—one by John F. Kennedy on 25 March 1961 and the second by Richard M. Nixon on 7 March 1970—were instrumental in providing the needed focus to America's space program. The Kennedy statement was crafted during a period of intense national introspection. The Soviet Union had just launched, and successfully recovered, the world's first cosmonaut. Though Yuri Gagarin spent just 89 minutes in orbit, his accomplishment electrified the world and caused the United States to question its scientific and engineering skills and its entire educational system. The American response—to land a man on the moon and return him safely to earth—bounded US space goals for the remainder of the decade. Prestige and international leadership were clearly the main objective of the Kennedy space program. However, the generous funding that accompanied the Apollo program had important residual benefits as well. It permitted the buildup of US space technology and the establishment of an across-the-board space capability to include planetary exploration, scientific endeavors, commercial applications, and military support systems.

However, as the decade of the 1960s drew to a close, a combination of factors, including domestic unrest, an upopular foreign war, and inflationary pressures, forced the nation to reassess the relative priority of the space program as compared to other national needs. It was against this backdrop that President Nixon made his long-awaited space policy statement in March 1970. It was a carefully considered and carefully worded statement that was clearly aware of political realities and the mood of Congress and the public. It said, in part:

> Space expenditures must take their proper place within a rigorous system of national priorities. . . .What we do in space from here on in must become a normal and regular part of our national life and must therefore be planned in conjunction with all of the other undertakings which are also important to us.*

Though spectacular lunar and planetary voyages continued until 1975, largely as a result of budgetary decisions made during the mid-1960s, it was clear that the Nixon administration considered the space program of intermediate priority and could not justify increased investment or the initiation of large new projects. It viewed space as a medium for exploiting and extending the technological and scientific gains that had already been realized. The emphasis was on practical space applications to include worldwide communications and meteorological systems, earth resource surveys, and scientific stellar and solar observations. Military surveillance satellites and navigations systems received increased emphasis as well.

One major new space initiative was undertaken during the 1970s. It was to have a far greater impact on the nation's space program than originally envisioned. The 1972 decision to proceed with development of a space shuttle was made for three principal reasons: a reusable vehicle promised to drastically reduce operational launch costs; the project would employ up to 40,000 aerospace workers—an important consideration in a presidential election year; and DOD expressed interest in the concept. It was inevitable that development problems would arise given the high technological risk associated with the program. When they did, and particularly when they began to manifest themselves in terms of cost overruns and schedule slippage, and to impact adversely already austere scientific and applications programs, it became apparent that no Americans would be launched into space for the remainder of the decade. The US space program had lost much of its early momentum.

Carter Administration Space Policy

Early in the Carter administration, a series of interdepartmental studies were conducted to address the malaise that had befallen the nation's space effort. These studies addressed apparent fragmentation and possible redundancy among the civil and national security sectors of the US space program and sought to develop a coherent recommendation for a new national space policy. The product of these efforts matured in the spring and fall of 1978, and resulted in two presidential directives (PD)—PD 37 on national space policy and PD 42 on civil space policy.

*Statement on the National Space Program, 7 March 1970

PD-37 reaffirmed the basic policy principles contained in the National Aeronautics and Space Act of 1958, and, for the first time, spelled out in coherent fashion the broad objectives of the US space program and the specific guidelines governing civil and national security space activities. The Reagan administration embodied most of these objectives and guidelines in its space policy. Therefore, we will discuss them in detail later.

PD-37 was important from a military perspective because it contained the initial, though tentative, indications that a shift was occurring in the national security establishment's views on space. Traditionally, they had viewed space as a force enhancer, that is, as a medium in which to deploy systems to increase the effectiveness of land, sea, and air forces. Although the focus of the Carter administration space policy was clearly on restricting the weaponization of space, PD 37 reflected an appreciation of the importance of space systems to national survival, a recognition of the Soviet threat to those systems, and a willingness to push ahead with development of an antisatellite (ASAT) capability in the absence of verifiable and comprehensive international agreements restricting such systems. In ot words, space was beginning to be viewed as a potential warfighting medium.

Whereas PD-37 addressed the space program in its totality, PD 42 was directed exclusively at the civil space community and was designed to set the direction for US efforts over the next decade. However, it was devoid of any long-term space goals, preferring instead to state that the nation would pursue a balanced evolutionary strategy of space applications, space science, and exploration activities. The absence of a more visionary policy reflected clearly the continuing developmental problems with the shuttle and the resulting commitment of larger than expected resources.

Reagan Administration Space Policy

National Security Decision Directive 42 (NSDD-42), publicly announced by President Reagan on 4 July 1982, embodies current national space policy. It supersedes all previous presidential space policy directives.

Basic goals. NSDD-42 reaffirms a national commitment to the exploration and use of space in support of the national well-being. It establishes the basic goals of US space policy as strengthening US security, maintaining US space leadership, obtaining economic and scientific benefits through space exploitation, expanding the investment and involvement of the US private sector in civil space and space-related activities, promoting international cooperative activities in the national interest, and cooperating with other nations in maintaining the freedom of space for activities that enhance the security and welfare of the entire human race.

Principles. NSDD 42 outlines broad principles that are to serve as the basis for the future US space program. For the most part, these principles have characterized, historically, the conduct of US space activities. These principles include five basic commitments. First, a commitment to the exploration and use of space by all nations for peaceful purposes and for the benefit of all people. For the first time, NSDD 42 defines peaceful purposes to permit activities in pursuit of national security goals. Second, there is a commitment to conduct international cooperative space-related activities that achieve scientific, political, economic, or national security benefits for the United States. Third, there is a commitment to pursue activities in space in support of the United States' inherent right of self-defense. Fourth, there is a commitment to develop Space Transportation System (STS) capabilities and capacities to meet appropriate national needs and to make the STS available to commercial and governmental users, both domestic and foreign. This includes the recognition that the STS is the primary space launch system for national security and civil government missions. Fifth, there is a commitment to continue to study space arms control options and to consider verifiable and equitable arms control measures that would ban, or otherwise limit, testing and deployment of specific weapons provided those measures are compatible with US national security.

Other broad principles include a rejection of national claims to sovereignty over any portion of space and celestial bodies as well as a rejection of any limitations on the fundamental right to acquire data from space. There is a recognition that space systems are national property and have the right to

pass through and operate in space without interference, and that the United States would view purposeful interference as an infringement on its sovereign rights.

These principles stress the encouragement of domestic commercial exploitation of space capabilities, technology, and systems for national economic benefit provided such exploitation is consistent with national security concerns, treaties, and international agreements. There is a mandate included that the national security and civil segments of the national space program closely coordinate, cooperate, and exchange information to avoid unnecessary duplication.

Specific policy guidance. In addition to outlining the basic goals and broad principles of the US space program, NSDD-42 provides specific policy guidance. The following is some specific guidance as it applies to military activities.

The United States will conduct activities in space deemed necessary to national security. These activities will support such functions as command and control, communication, navigation, environmental monitoring, warning, surveillance, and self-defense.

The United States will pursue survivability and endurance of space systems, including all system elements, commensurate with their planned use in crisis and conflict, with the threat, and with the availability of other assets to perform the mission. Deficiencies are to be identified and eliminated. The United States will undertake a long-term program to provide more assured survivability and endurance.

Development of an antisatellite (ASAT) capability is to proceed with a goal of operational deployment. The primary purposes of a US ASAT capability are to deter threats to the United States and allied space systems and within limits of international law, it is to deny any adversary the use of space-based systems that support hostile military forces.

Another specific policy guidance concerns attack warning. It states that the United States will develop and maintain an integrated attack warning, notification, verification, and contingency reaction capability that can detect and react effectively to threats to US space systems. Unique national security considerations may require developing special purpose launch capabilities.

Several policy guidelines concerning the STS have been established. The STS is to be afforded the degree of survivability and security protection required for a critical national resource. The first priority of the STS program is to make the system fully operational and cost effective in providing routine access to space. However, national security missions will receive STS launch priority with the DOD in operational control of national security missions. Enhancement of STS operational capability, upper stages, and methods of deploying and retrieving payloads is to be pursued as national requirements are defined. The design of US government spacecraft will be such that they can take advantage of the unique capabilities of the STS. Transition to the shuttle is to be completed as expeditiously as practical.

The single statement of national policy that could most influence the future direction of military activities in space and that clearly reflects the transition to a space warfighting perspective, resides in NSDD 85, dated 25 March 1983. In this document, President Reagan states as his long-term goal the elimination of the threat presently posed by nuclear ballistic missiles. This is an objective he hopes to achieve by reducing US reliance on the threat of retaliation by strategic nuclear forces and by simultaneously increasing the contribution of strategic defensive forces. Though specific systems are not identified, it is recognized generally that space-based systems will play a significant role in this vision in the future.

Department of Defense Space Policy

DOD directive 5100.1, dated 31 December 1958, provides that the "Department of Defense shall maintain and employ armed forces . . . to insure, by timely and effective military action, the security of the United States, its possessions, and areas vital to its interests (and) to uphold and advance the national policies and interests of the United States." The directive does not explicitly mention space. However, neither does it exclude the use of space to further national security. Therefore, together with the National Aeronautics and Space Act, this directive provides a solid foundation for the military use of space.

The most recent statement of DOD space policy occurred on 22 June 1982. Though released prior to NSDD-42, the DOD policy is fully consistent with and supports the principles contained in the national space policy. Table 15-2 summarizes the DOD space policy and shows how it relates to and seeks to implement national policy.

Air Force Space Policy

The earliest recorded statement of Air Force policy regarding space occurred on 15 January 1948, when Gen Hoyt S. Vandenberg stated "The USAF, as the service dealing primarily with air weapons—especially strategic—has logical responsibility for the satellite." As reflected in General Vandenberg's statement, traditionally, Air Force leaders have viewed space as a medium in which the Air Force would have principal mission responsibilities. This view was perhaps best articulated by former Air Force Chief of Staff Gen Thomas D. White, when he coined the term "aerospace" in the following testimony before the House Committee on Science and Astronautics in February 1959:

> Since there is no dividing line, no natural barrier separating these two areas (air and space), there can be no operational boundary between them. Thus air and space comprise a single continuous operational field in which the Air Force must continue to function. The area is aerospace.

Though the Air Force had staked its claim to inherent rights in space, it was possibly the least capable of the military services during the late-1950s to accomplish the space mission. The government had given the Navy the responsibility for launching America's first satellite. This was Project Vanguard. After several disheartening attempts, they placed a satellite in orbit in March 1958. The Army had launched successfully the first US space system, known as Explorer I, in January 1958, and had assembled an impressive array of scientific and engineering talent under the leadership of Dr Werhner von Braun to work on a new, more capable launch system—the Saturn booster. By contrast, the Air Force had not attempted the launch of even a single space system.

However, the years from 1958 through 1961 were watershed years for the Air Force. The creation of NASA in 1958 prompted the transfer of a number of scientific space programs from the military services, primarily from the Army and Navy, to the new civilian agency. The late-1959 decision by the secretary of defense to assign responsibility to the Air Force for booster development and payload integration provided the foundation for the emerging Air Force space program and indirectly resulted in the transfer of the Army's Saturn program to NASA. However, the key decision was made in early 1961 when the Secretary of Defense McNamara, in an effort to eliminate duplication and competition among the services, assigned all space research and development projects to the Air Force. The only exceptions were the Navy's Transit navigation satellite system and the Army's Advent communications satellite program. Though it was subsequently to be modified in 1970 to permit assignment of responsibility for space programs on a case-by-case basis, this 1961 decision catapulted the Air Force into a leadership position in the development, acquisition, operation, and use of military space systems.

No single document prescribes Air Force space policy. However, a collection of documents, developed at various times in response to the growing US dependence on space systems and the growing Soviet threat to the free use of space, provides broad guidance to Air Force activities in developing plans, requirements, and programs for space systems. These documents include two chief of staff letters, dated 9 May 1977 and 1 July 1983, and the March 1983 *Air Force Space Plan*.

These documents acknowledge that the Air Force has exclusive responsibilities in space in certain areas. These responsibilities include the protection of the right to free use of space by providing space superiority; the responsibility for all other DOD space systems vital to performing Air Force missions; the development of space-based weapons if such weapons respond to national security requirements; and the responsibility for all DOD space support activities encompassing launch, on-orbit control, and space systems acquisition. The document acknowledges that, as the DOD executive agent for liaison with NASA, the Air Force accepts its responsibilities to closely coordinate and cooperate on projects of mutual interest.

Table 15-2.
DOD Implementation of National Space Policy

Subject	*NSDD-42*	*DOD Space Policy*
Space Operational Force Support	Conduct activities in space necessary to national security	Provide effective operational support to US forces in peacetime and conflict. Future use of space should have an operational orientation.
	Support following functions: Command and Control Communications Navigation Environmental Monitoring Warning Surveillance Space Defense	Support following functions: Command and Control Communications Navigation Environmental Monitoring Warning Surveillance Space Defense
	Develop and maintain an integrated attack warning, notification, verification, and contingency reaction capability to effectively detect and react to threats to US space systems.	
	Maintain and coordinate separate national security and civil operational space systems when requirements dictate.	
	Closely coordinate national security and civil space programs and emphasize technology sharing within necessary security constraints.	
Space Launch	The Space Transportation System (STS) is the primary launch system for both national security and civil missions.	Recognizes the STS as the primary space launch system.
	Design spacecraft to take advantage of unique STS capabilities.	
	Complete transition to STS as expeditiously as practical.	
	Continue Expendable Launch Vehicle (ELV) Operations until STS capabilities can meet needs and obligations.	
	Unique requirements may dictate development of special purpose launch capabilities.	Requires the availability of an adequate launch capability to provide flexible and responsive access to space to meet national security requirements.
	Pursue enhancements as national requirements are defined.	
	Afford STS the degree of survivability and security protection required for a critical national space resource.	

Table 15-2 (Continued).
DOD Implementation of National Space Policy

Subject	*NSDD-42*	*DOD Space Policy*
STS Operations & Management	NASA will continue to develop the STS in cooperation with DOD. DOD will integrate national security missions into STS and exercise operational control over those missions. Launch priority provided for national security missions.	Continue to cooperate with NASA to develop a fully operational STS. Plan and conduct all activities necessary to control and manage national security STS missions.
Antisatellite System (ASAT)	Develop an ASAT capability with a goal of operational deployment. Purpose is to deter threats to US and Allied space systems and, within limits imposed by international law, deny an adversary the use of space-based systems to support military forces.	Develop an operational ASAT capability within limits imposed by international law. Purpose to deter hostile threats to US and allied space systems and threats from space systems supporting hostile military forces.
Survivability & Endurance	Pursue survivability and endurance of space systems, including all system elements, commensurate with planned use in crisis and conflict, the threat, and availability of other assets to perform mission. Identify and eliminate deficiencies. Undertake an aggressive, long-term program to provide more assured survivability and endurance.	Design, develop, and operate military space systems, to include all essential ground elements, to enhance survivability and endurance of critical mission functions.
Technology	Maintain US space leadership	Direct continued maintenance of a strong technology base. Emphasize leadership in those technology areas necessary for effective defense.
Arms Control	Consider verifiable and equitable arms control measures that would ban or otherwise limit testing and deployment of specific weapons systems provided those measures are compatible with US national security.	Emphasize continued adherence to the existing international space legal regime. Consider verifiable and equitable arms control measures that would ban or otherwise limit the deployment of specific weapons systems provided those measures are compatible with US national security.

Of the three policy documents, the most comprehensive is the Air Force Space Plan. It does three things. It provides broad guidance affecting future Air Force investment and management strategy for space activities; provides a vehicle to implement national and DOD space policies; and identifies, and resolves, where possible, issues affecting Air Force use of the space environment. Though containing details that require the document to be classified, the basic thrust of the space plan is unclassified, that is, that space is a place, not a mission. Therefore, as with the other environments of the land, sea, and air, the systems deployed in space have the potential to perform combat or combat support missions. We can divide these missions into four broad categories: space control, force application, force enhancement, and space support.

Space control encompasses those combat missions that provide freedom of action in space for friendly forces while denying it to an enemy. It embodies the concept of space superiority and consists of two parts, counterspace operations and space interdiction. Counterspace operations are those spaceborne or terrestrial operations conducted to gain control of and dominance over the space medium. Included are measures to ensure that friendly space forces have freedom of action through space. Space interdiction includes operations conducted against the enemy's space lines of communication (space systems used to support or participate in military operations) that could be used to support operations against friendly forces.

Force application is concerned with combat missions conducted from space against terrestrial targets, primarily. The objective would be the influencing of a conflict. Force application from space could potentially encompass traditional Air Force combat missions such as strategic offense and defense, interdiction of enemy forces, and close support, among others.

Force enhancement includes combat support missions involving the use of space systems to improve the effectiveness of surface, sea, air, and space forces. It includes such support functions as communications; surveillance (that is, tactical warning and meteorology); reconnaissance; navigation and positioning; mapping, charting, and geodesy; and search and rescue aids.

The final broad category is space support. This encompasses combat support missions involving all necessary prelaunch preparations, as well as the activities involved with deploying and sustaining space systems. It includes such functions as space launch and recovery (the preparation, buildup, launch, deployment, and if necessary, retrieval of space systems as well as the space transporter); and orbit transfer (operations involving the use of propulsive stages to maneuver satellites from their initial orbit to their final mission orbit). It includes on-orbit control (operations to plan, direct, and sustain deployed military space systems) and management, planning, and operations support activities (manpower, training, safety, education, and logistics support).

Historically, military space power has concentrated on combat support missions, specifically force enhancement-type missions. Systems performing functions such as communications, environmental observation, photoreconnaissance, tactical warning, and navigation dominate the military space program. The dilemma that presently confronts military space planners is that these systems have become so capable and in some instances so important to the conduct of terrestrial military operations that their unhindered use could prove decisive in future conflicts. Therefore, we are likely to see a change in mission emphasis in the coming years. The space control role will probably be increasingly pursued as a means to ensure that US and allied forces can operate freely in space while being in a position to deny the same freedom to an enemy. In addition, technology now enables space systems to complement traditional methods of bringing force to bear upon an enemy. The future is likely to witness a more balanced mixture of the four space roles.

AIR FORCE SPACE DOCTRINE

Though the government had articulated national, DOD, and Air Force space policies in one format or another for better than two decades, the Air Force did not have a space doctrine until it published AFM 1 6, *Aerospace Basic Doctrine, Military Space Doctrine*, on 15 October 1982. We can attribute the long delay, at least partially, to the nature of doctrine and the influence of history on its

development. JCS Publication 1 defines doctrine as "fundamental principles by which the military forces guide their actions in support of objectives. It is authoritative but requires judgment in application." Professor I. B. Holley, Jr, of Duke University, provided a shorter and perhaps more workable definition when he stated that "military doctrine is what is officially believed and taught about the best way to conduct military affairs." Gen Thomas D. White stated that doctrine is based partially on history but that it is also very dynamic.

> In the development of superior air leadership, the education process cannot treat air doctrine as a set of abstract principles to be learned by rote, like mathematical formulas, and dutifully filed away for future reference. Air Force doctrine is made up, not of abstraction, but of dynamic living truths forged in the heat of combat and tested in the crucible of war.

Historical experience influences official doctrine and anticipated developments in warfighting shape it. The fact that military experience in space has been limited and the absence of any "dynamic living truths forged in the heat of combat" undoubtedly help to explain the lapse of 20 years before the Air Force published its first official space doctrine.

Though published prior to the *Air Force Space Plan* and the July 1983 chief of staff statement on Air Force space policy, AFM 1-6 nevertheless clearly reflects the changing emphasis in the military use of space. The authors recognized the inherent benefits to be gained by any nation that chooses to exploit the military advantages of space. They specifically chartered the Air Force "to provide forces for controlling space operations and gaining and maintaining space superiority." These are missions essential to preserve free access to, and transmit through, space for peaceful purposes for both the military and the civil sectors.

AFM 1-6 summarizes those principles that form the basis for Air Force responsibilities in space. These principles include preserving space for peaceful purposes by providing appropriate military force for deterrent operations in space. This military force must be ensured by developing space systems that support national security objectives and integrating those systems into the military structure to increase effectiveness, readiness, reliability, survivability, and sustainability of warfighting systems. At the same time, the Air Force must ensure that the goals for military space use are clearly defined and understood and that close coordination and cooperation with NASA and the civil sector on space projects of mutual interest and benefit are carried out.

These principles lead to the creation of many Air Force responsibilities in space. In the area of development, the Air Force is responsible for developing forces for conducting military operations within the unified or specified command structure and providing operational concepts and employment tactics to integrate space system support into these unified and specified commands. This includes the developing of the plans, programs, policy, and implementation and employment strategies to integrate space forces into the US military posture. The Air Force must develop, operate, sustain, and employ space systems, including increasing the effectiveness, readiness, and survivability of weapon systems, to ensure the capability to deter and resolve conflict on terms favorable to the United States. It must ensure the needed education and training programs for individuals directing space forces as well.

In the area of management, the Air Force is responsible for managing military space operations. This includes the launch, command and control, on-orbit sustainment, and the refurbishment of military space needs throughout the DOD, NASA, and other governmental departments and branches.

The Air Force must sustain the potential for military operations by applying superior space-related technologies. This would include capabilities for collecting, processing, and disseminating information as well as planning for deployment and employment of military space forces. The Air Force will accomplish this by encouraging innovation to take advantage of advances in science and technology; developing and maintaining capability to provide research and development, testing, engineering, and life cycle support for space systems; developing the technology base and research, development, and acquisition policies that accommodate procurement requirements for space systems; and providing the acquisition service authorized by DOD as well as other executive services. The Air

Force must encourage innovation to take advantage of advances in technology in all of these areas.

The Air Force is charged with providing space-based surveillance, warning, and attack assessment to alert national command authorities and military commanders of impending or potential attacks against the United States, its allies, or their forces. This extends to include negating enemy attacks to, from, in, or through space.

Finally, the Air Force is responsible for ensuring that space is considered as a medium from which Air Force missions and tasks potentially can be accomplished. This will be done by ensuring that space programs are clearly understood and fiscally attainable; providing requirements and advocacy for military space systems; and providing programming data, operational strategies, and requirements information to justify financial support for military space programs.

With respect to potential warfighting missions, AFM 1-6 contains this statement: "Space-based weapon systems could contribute to deterrence in peacetime and to more rapid conflict termination or increased survivability in war." Although it acknowledges potential warfighting missions, AFM 1-6 further states that employment of military space systems will depend on such operation considerations as reliability, survivability, security, and flexibility; their effectiveness compared to alternative solutions; and the impact of combined surface and space operations on national security objectives.

In looking to the future, AFM 1-6 specifies four areas of growth in military space activities.

- Expansion of space support operations with on-orbit resources that include space stations, shuttle, orbital transfer vehicles, energy generators, and manufacturing processes.

- Development and continued evolution of quick reaction launch capabilities with short turnaround from more survivable launch facilities.

- Development of the capability to support multipurpose operations in space (such operations could include space-to-space, space-to-earth, and earth-to-space activities).

- Development of space systems for rapid termination of any military conflict.

SUMMARY

The responsibilities of the Air Force in space include a large and growing number of functions that contribute to the defense of the United States. Space operations are important elements of a credible deterrent to armed conflict. In the near-term, should deterrence fail, space operations will help resolve a conflict on acceptable terms to the United States by providing various kinds of information and support to military forces. In the future, it is possible that space systems will provide the decisive edge in countering threats to our national interests.

The Air Force regards military operations in space among its prime responsibilities and conducts these operations according to the letter and spirit of existing treaties and international law. In response to national direction, it must ensure freedom of access to space for peaceful pursuits. In carrying out aerospace missions for defense of the United States, it uses space systems to perform unique, economical, and effective functions to enhance the nation's land, sea, and air forces. It serves client relationships in the management of space programs and meets user priorities established by the Department of Defense. In view of the growing threat to the free use of space and the critical role of space operations in the defense of the United States, the Air Force has moved toward increased institutionalization of military space missions so that the Department of Defense can draw clear lines of responsibility for accomplishing these missions.

BIBLIOGRAPHY

Arms Control and Disarmament Agency. *Arms Control and Disarmament Agreements.* Arms Control and Disarmament Agency, Washington, D.C., 1982.

Bridge, Robert L., Lt Col. "International Law and Military Activities in Outer Space." *Akron Law Review* 13, no. 4, Spring 1980:649-664.

Butler, A. J. Lt Col. "Peaceful Use and Self-Defense in Outer Space." *Proceedings of the 1983 Symposium on Military Space Communications and Operations*. USAF Academy, 2 4 August 1983:87 92.

Convention on Registration of Objects Launched into Outer Space. 28 UST 695, TIAS no. 8480, 14 January 1978.

Department of Defense. *Defense Guidance, FY 1986-1990*. Department of Defense, Washington, D.C., 2 March 1984.

Department of Defense. *Department of Defense Space Policy*. Washington, D.C., 22 June 1982.

Department of Defense. *DOD Directive 1500.1* as Amended. Washington, D.C., 31 December 1958.

Department of the Air Force. *United States Air Force Space Plan*. Department of the Air Force, Washington, D.C., March 1983.

Futrell, Robert F. *Ideas, Concepts, Doctrine: A History of Basic Thinking in the United States Air Force, 1907-1964*. Maxwell Air Force Base, Alabama, Air University, 1971.

Gabriel, Charles A., Gen. *Air Force Space Roles and Responsibilities*. Letter to all major commands and special agency commanders, 1 July 1983.

Hungerford, John B., Jr, Maj. "Organization for Military Space A Historical Perspective." Air Command and Staff College report number 82-1235, Maxwell Air Force Base, Alabama, Air University, March 1982.

National Aeronautics and Space Act of 1958, as Amended. Public Law 85-568, 72 Stat. 426, 29 July 1958.

National Security Decision Directive Number 42, National Space Policy. White House, Washington, D.C., 4 July 1982.

Office of Technology Assessment. *Civilian Space Policy and Applications*. Office of Technology Assessment, Washington D.C., June 1982.

Office of the White House Press Secretary. *White House Fact Sheet, US Civil Space Policy*. Office of the White House Press Secretary, Washington, D.C., 11 October 1978.

United Nations Charter, 26 June 1945.

United States Air Force. *Aerospace Basic Doctrine, Military Space Doctrine*. AFM 1 6. Department of the Air Force, Washington, D.C., 15 October 1982.

United States Air Force Historical Division Liaison Office. *The Threshold of Space—The Air Force in the National Space Program, 1945-1959*. United States Air Force Historical Division Liaison Office, Washington, D.C., September 1960.

Appendix A

THE DEORBITING PROBLEM

The general operation of moving a body from an earth orbit to a precise point on the surface of the earth is a difficult problem. The orbit may be circular or elliptical, high or low. It may be inclined at various degrees to the rotational axis of the earth. The deorbiting maneuver may be the lofted, depressed, or retroapproach. This retroapproach is the only method discussed in this section. Time of flight is a major parameter in deorbiting and presents one of the more difficult mathematical problems. However, the theory developed here is sufficiently general in scope that it may be modified to solve most problems. The following simplified illustration reduces the complexity of the problem. The illustration discusses deorbiting from a circular polar orbit from a point over the North Pole by assuming no atmosphere.* The approach and solution will permit impacting any earth target provided fuel for retrothrust is no limitation.

Consider a target at 60° north latitude. If we could stop the earth's rotation (see fig. A-1) with the target in the plane of the circular orbit, then the only necessity would be to apply retrovelocity (Δv_1) to the orbital body so it would traverse the ellipse shown (arc 1-4) and impact the target. The reentry velocity magnitude (v_b), equal to the circular speed (v_c) minus the retrovelocity (Δv_1), would be necessary at apogee (1) to impact any target on the 60° north parallel. But, the direction of v_b would have to be oriented with respect to the earth's axis so that it would lie in the plane formed by the earth's axis and the predicted position of the target.

Since the earth is rotating at a constant speed, we must know the time of flight (t_b) of the orbital object to predict where the target will be at impact. For example, when the orbital vehicle is at point 1 over the North Pole, the target is at point 2. However, during the descent of the payload, the target moves to point 3. To predict point three, we must know t_b.

If we know the magnitudes of v_c, v_b, and α, the angle between these two vectors, we can use the law of cosines to determine Δv_2. Application of Δv_2 to v_c ensures that v_b will have the proper magnitude and direction so that the object will impact a selected target on a rotating earth.

DEORBITING VELOCITY

Looking first at the geometry of the problem (see fig. A-2), notice that the total problem lies in one plane. Basically, we know the radius of the circular orbit (r_c) and the latitude of the target (L). Since the original circular orbit and the bombing transfer ellipse are coplanar and cotangential, the problem begins as a Hohmann transfer. To determine Δv, we can compute the rearward velocity increment that must be applied to cause the object to impact the target from $\Delta v = v_c - v_b$, if we know v_c and v_b.

$$v_c = \sqrt{\frac{\mu}{r_c}}$$

$$v_b = \sqrt{\frac{2\mu}{r_a} - \frac{\mu}{a}}$$

*The reader may consider the effect of the atmosphere by using the radius of reentry rather than the radius of the earth, and by then considering the range and time of flight for a specific reentry body. The solution may be modified to consider elliptical orbits and inclined orbits.

$$\text{Since } r_a = a + c = a(1+\epsilon),\ a_b = \frac{r_a}{1+\epsilon_b}$$

$$\text{Then } v_b = \sqrt{\frac{2\mu}{r_a} - \mu \frac{(1+\epsilon_b)}{r_a}} = \sqrt{\frac{\mu(1-\epsilon_b)}{r_a}}$$

Write the general equation of the conic as $k\epsilon = r(1 + \epsilon \cos \nu)$ and evaluate at the two known points, r_e* and r_a, noting that at impact $\cos \nu = -\cos \theta_t = -\sin L$, and at retrofire $\cos(180°) = -1$.

Then $k\epsilon = r_e(1 - \epsilon_b \sin L) = r_a(1 - \epsilon_b)$

Solving for ϵ_b.

$$\epsilon_b(r_a - r_e \sin L) = r_a - r_e$$

$$\epsilon = \frac{r_a - r_e}{r_a - r_e \sin L}\ .$$

To find ϵ_b, and in turn, the magnitude of v_b, we need to know only the latitude of the target and the attitude circular orbit. Suppose there is a satellite in a 500 nautical mile circular polar orbit. When the orbital vehicle arrives directly over the North Pole, it is desired to deorbit an object that will impact at 60° north latitude. Assuming that the earth has no atmosphere and is nonrotating, what retrovelocity is necessary?

$$\epsilon_b = \frac{r_a - r_e}{r_a - r_e \sin L} = \frac{(23.94 \times 10^6 \text{ ft}) - (20.9 \times 10^6 \text{ ft})}{(23.94 \times 10^6 \text{ ft}) - (20.9 \times 10^6 \text{ ft})} \quad \sin 60°$$

$$\epsilon_b = .520$$

$$v_b = \sqrt{\frac{\mu}{r_a}(1 - \epsilon_b)} \quad \sqrt{\frac{14.08 \times 10^{15} \text{ ft}^3/\text{sec}^2\ (1 - .520)}{23.94 \times 10^6 \text{ ft}}}$$

$$v_b = 16{,}790 \text{ ft/sec}$$

$$\Delta v = v_c - v_b = 24{,}240 \text{ ft/sec} - 16{,}790 \text{ ft/sec}$$

$$\Delta v = 7{,}450 \text{ ft/sec}\ .$$

Looking again at the geometry, note that there is a satellite in a 500 nautical mile circular, polar orbit with $v_c = 24{,}240$ ft/sec. A retrovelociy increment, Δv, equal to 7,450 ft/sec was applied. This provided a magnitude $v_b = 16{,}790$ ft/sec so that now the object follows arc 1-2 and impacts on 60° north latitude.

Figure A-3 provides a graph of velocity versus latitude to determine the magnitude of v_b. The graph is for a specific altitude, for release over the North pole, and for no atmosphere. Included are more general graphs in figures A-4 and A-5.

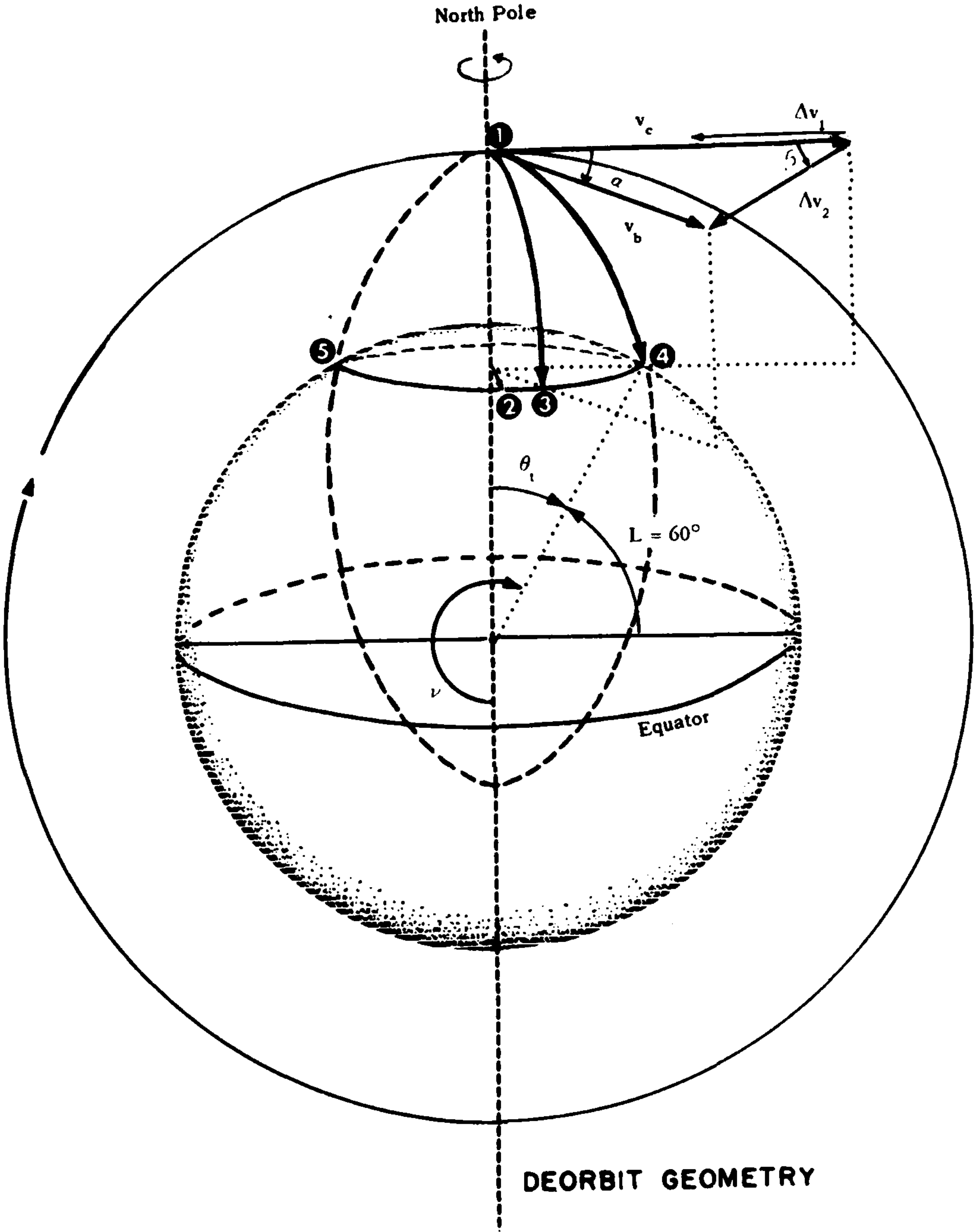
DEORBIT TIME OF FLIGHT

As the bomb falls from apogee, the target moves toward the east due to earth rotation. Its speed is 1,520 cos L ft/sec, so the necessity to compute accurately the time to bomb, t_b, is readily apparent.

First, look at the problem in schematic figure A-1. Recall that at the time the vehicle is over the North Pole at point 1, the target is at point 2. Thus, we must know t_b so that we can predict the location of point 3. If we know the meridian that point 3 will be on, and the meridian with which the circular orbit coincides at the instant of deorbit, then we can determine the angle, α, by subtraction.

There are several methods for computing t_b from release to impact. The general time of flight method is presented here.

*r_e is radius of the earth, but radius of reentry could be used if desired.



North Pole
v_c
Δv_1
Δv_2
v_b
α
β
1
2
3
4
5
θ_t
L = 60°
ν
Equator
DEORBIT GEOMETRY

Figure A-1.

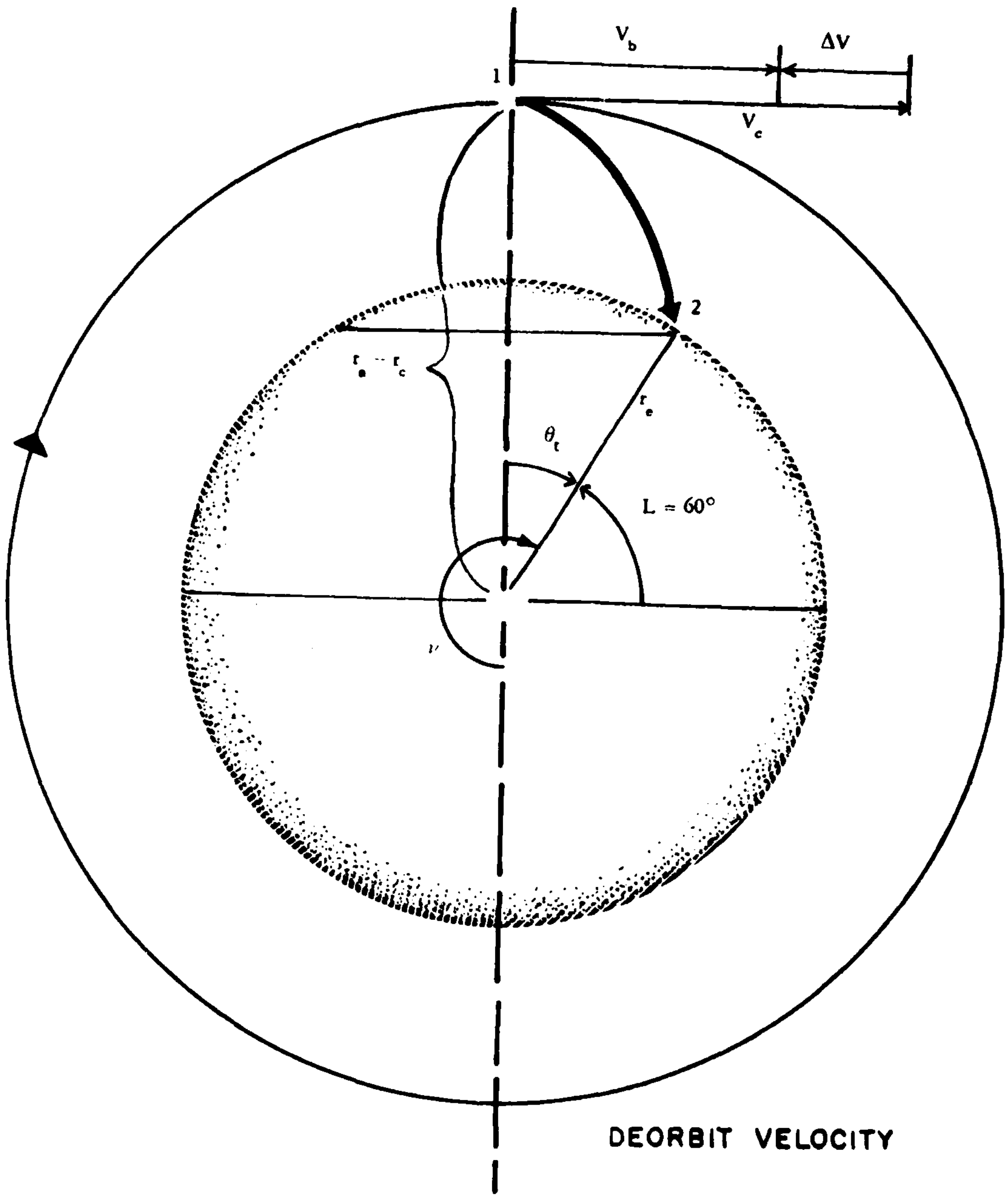


Figure A-2.

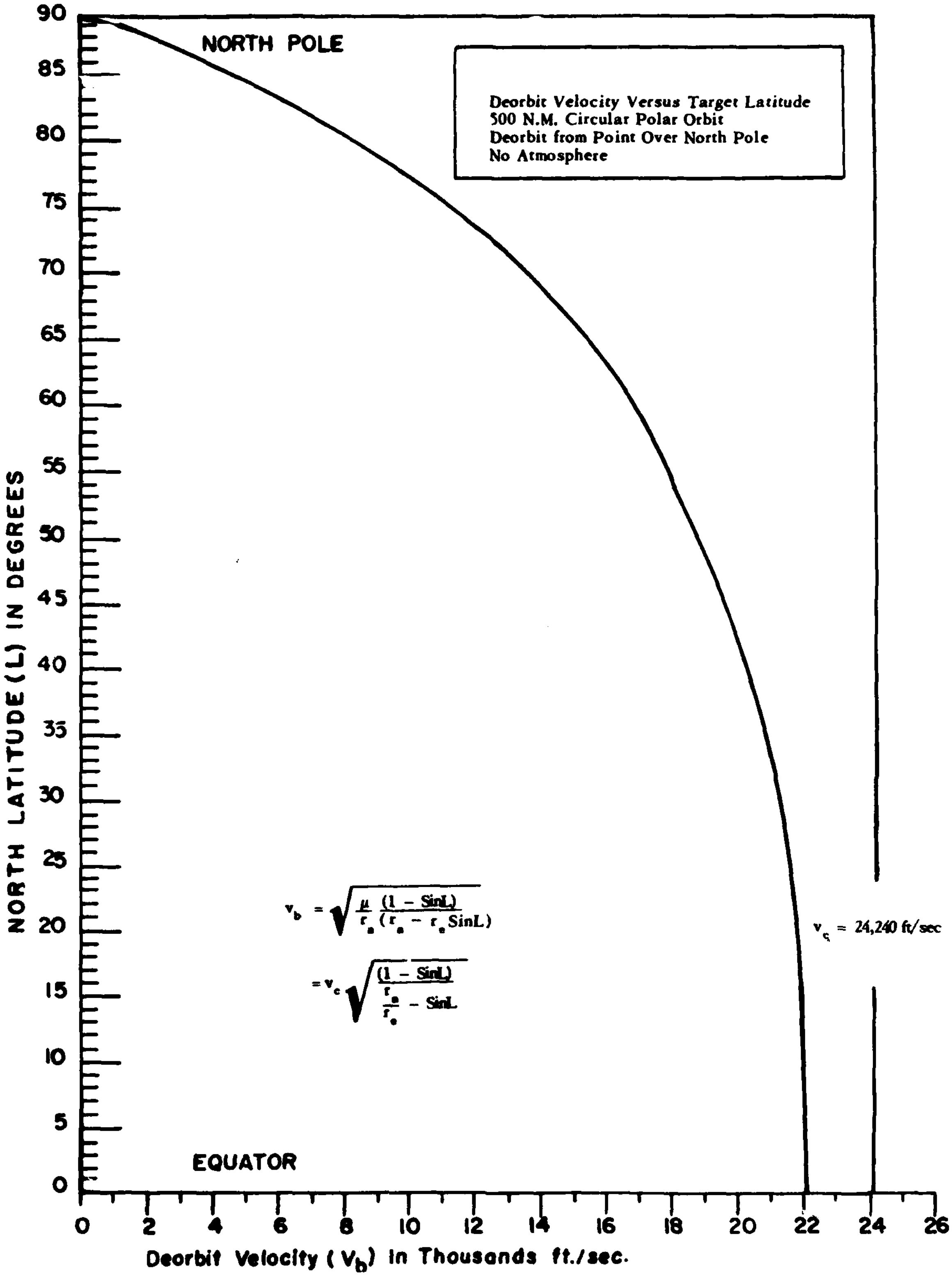

NORTH POLE
Deorbit Velocity Versus Target Latitude
500 N.M. Circular Polar Orbit
Deorbit from Point Over North Pole
No Atmosphere
EQUATOR
NORTH LATITUDE (L) IN DEGREES
Deorbit Velocity (V_b) in Thousands ft./sec.
$v_b = \sqrt{\frac{\mu}{r_a}\frac{(1 - \text{Sin}L)}{(r_a - r_e \text{Sin}L)}}$
$= v_c\sqrt{\frac{(1 - \text{Sin}L)}{\frac{r_a}{r_e} - \text{Sin}L}}$
v_c = 24,240 ft/sec
0 2 4 6 8 10 12 14 16 18 20 22 24 26
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90

Figure A-3.

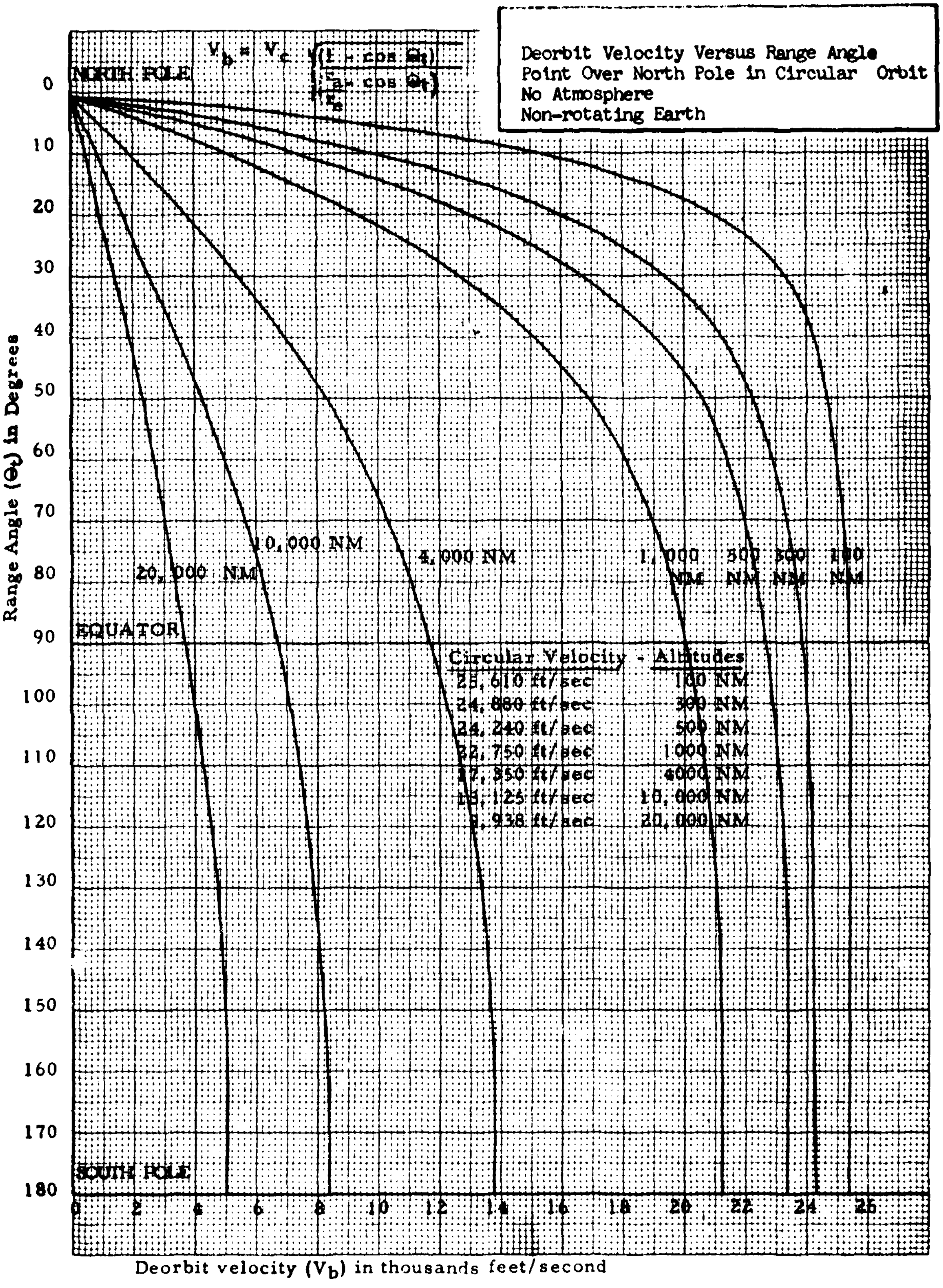
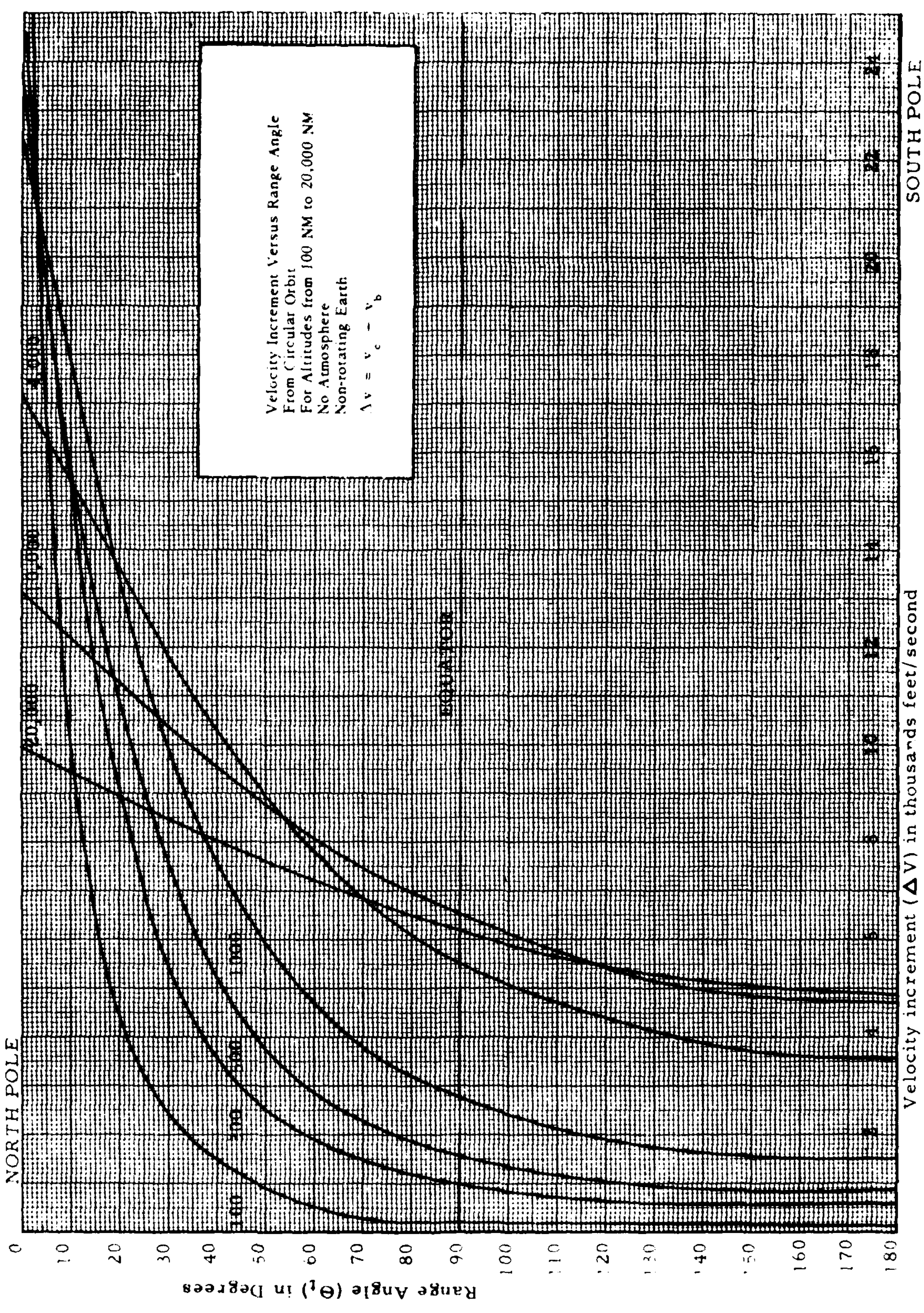

Deorbit Velocity Versus Range Angle
Point Over North Pole in Circular Orbit
No Atmosphere
Non-rotating Earth
$V_b = V_c \sqrt{\frac{(1 - \cos \Theta_t)}{(\frac{r_a}{r_e} - \cos \Theta_t)}}$
NORTH POLE
EQUATOR
SOUTH POLE
20,000 NM
10,000 NM
4,000 NM
1,000 NM
500 NM
300 NM
100 NM
Circular Velocity - Altitudes
25,610 ft/sec 100 NM
24,880 ft/sec 300 NM
24,240 ft/sec 500 NM
22,750 ft/sec 1000 NM
17,350 ft/sec 4000 NM
13,125 ft/sec 10,000 NM
9,938 ft/sec 20,000 NM
Range Angle (Θ_t) in Degrees
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
0
2
4
6
8
10
12
14
16
18
20
22
24
26
Deorbit velocity (V_b) in thousands feet/second

Figure A-4



Velocity Increment Versus Range Angle
From Circular Orbit
For Altitudes from 100 NM to 20,000 NM
No Atmosphere
Non-rotating Earth
$\Delta v = v_c - v_b$
NORTH POLE
SOUTH POLE
EQUATOR
100
200
500
1,000
20,000
Velocity increment (ΔV) in thousands feet/second
Range Angle (Θ_1) in Degrees

Figure A-5

$$t_{1 \rightarrow 2} = \sqrt{\frac{a^3}{}}(u_2 - \epsilon \sin u_2) - (u_1 - \epsilon \sin u_1)$$

Since position 1 corresponds to apogee:

$$u_1 = \pi, \sin u_1 = 0$$

Then $t_b = t_{1 \rightarrow 2} = \sqrt{\frac{a^3}{\mu}}(u_t - \epsilon_\rho \sin u_t - \pi)$

$$a = \frac{r_a}{1 + \epsilon_b}$$

$$\cos u_t = \frac{\epsilon_b - \cos\theta_t}{1 - \epsilon \cos \theta_t} = \frac{\epsilon_b - \sin L}{1 - \epsilon_b \sin L}$$

All of these parameters are familiar except u, which is the eccentric anomaly (see fig. A-6). If a perpendicular is dropped through the target to the major axis of the ellipse, it intersects a circle (with the center at C, and the diameter equal to the major axis of the ellipse) at point Q. By definition, angle BCQ is u, the eccentric anomaly. The radius of the circular orbit, r_a, is given. We can calculate v_b and ϵ_b from formulas previously given.

Working with the same example that we used to illustrate deorbiting velocity, we will tabulate t_b from only two known quantities, the altitude of the satellite and the latitude of the target. A satellite is in a 500 nautical mile circular polar orbit. Find the time of flight, t_b, from directly over the North Pole to a target on the 60° north parallel:

$$\epsilon_b = \frac{r_a - r_e}{r_a - r_e \sin L} = \frac{(23.94 \times 10^6 \text{ ft}) - (20.9 \times 10^6 \text{ ft})}{(23.94 \times 10^6 \text{ ft}) - (20.9\ 10^6 \text{ ft}) \sin 60°}$$

$$\epsilon_b = .520$$

$$v_b = \sqrt{\frac{\mu}{r_a}(1 - \epsilon_b)} = \sqrt{\frac{14.08 \times 10^{15} \text{ ft}^3/\text{sec}^2\ (1 - .520)}{23.94 \times 10^6 \text{ ft}}}$$

$$v_b = 16{,}790 \text{ ft/sec}$$

$$a = \frac{r_a}{1 + \epsilon_b} = \frac{23.94 \times 10^6 \text{ ft}}{1.520} = 15.78 \times 10^6 \text{ ft}$$

$$\cos u_t = \frac{\epsilon_b - \sin L}{1 - \epsilon_b \sin L} = \frac{.520 - \sin 60°}{1 - .520 \sin 60°} = -.637$$

Noting u_t lies between 180° and 270°:

$$u_t = 360° - 129.6° = 230.4° = 4.084 \text{ radians}$$

$$t_b = \sqrt{\frac{a^3}{\mu}}[u_t - \epsilon_b \sin u_t - \pi]$$

$$t_b = \sqrt{\frac{(15.78 \times 10^6 \text{ ft})^3}{14.08 \times 10^{15} \text{ ft}^3/\text{sec}^2}}[4.084 - .520 \sin 230.4° - 3.1416]$$

$$t_b = 676 \text{ sec} = 11.27 \text{ min}$$

Figure A-7 shows t_b in minutes versus a plot of target latitude. Once again we must recognize that this chart is for a specific orbital altitude and no atmosphere. See also the more general graph in figure A-8.

Consider impacting a target at 60° north latitude and 30° east longitude. The plane of the satellite is coincident with the 70° east meridian. This means that, when the satellite is at point one and the target is at point two, the angle between the target and orbital plane is 40°. However, the target is moving. It moves at:

TIME OF FLIGHT GEOMETRY

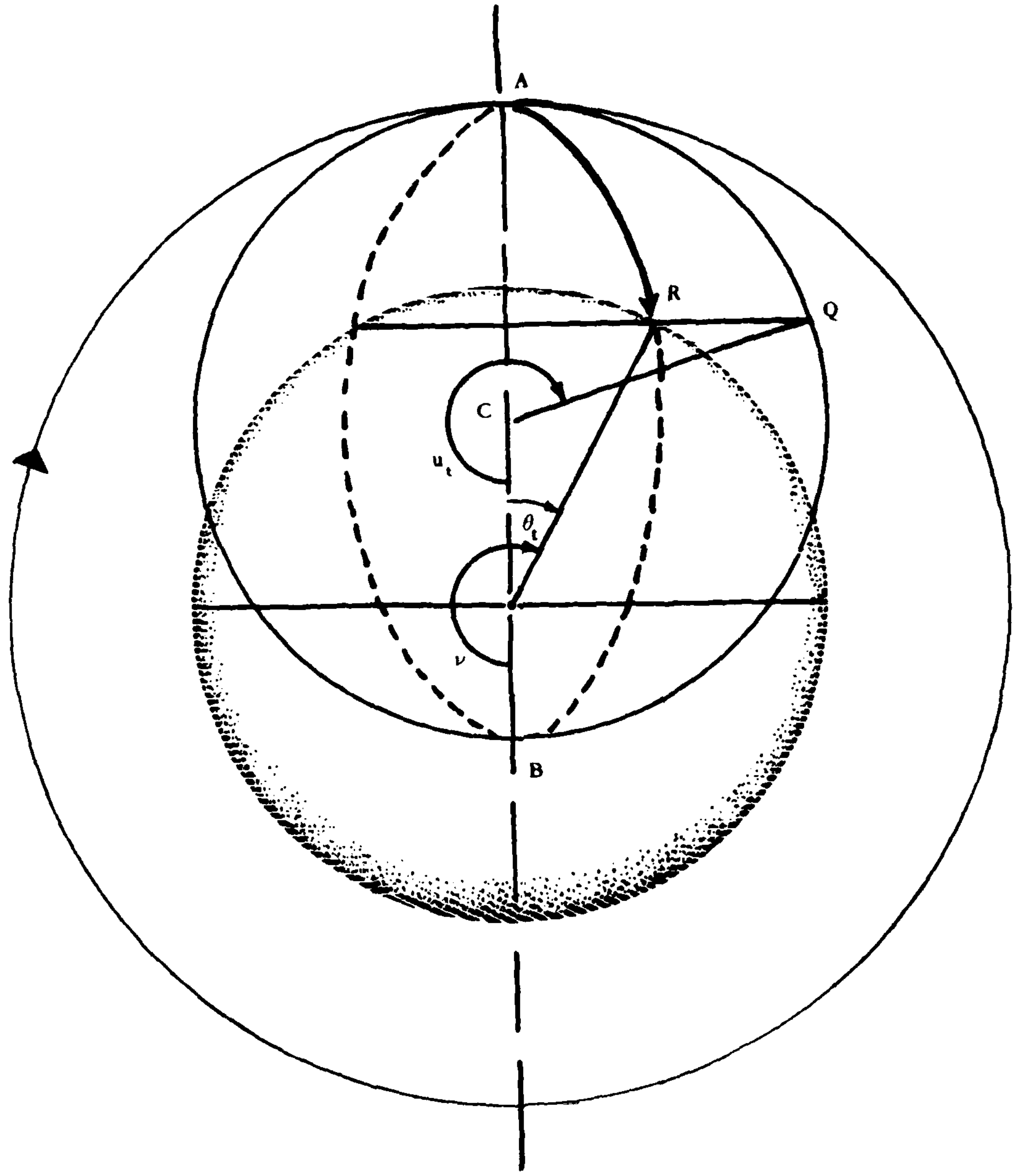
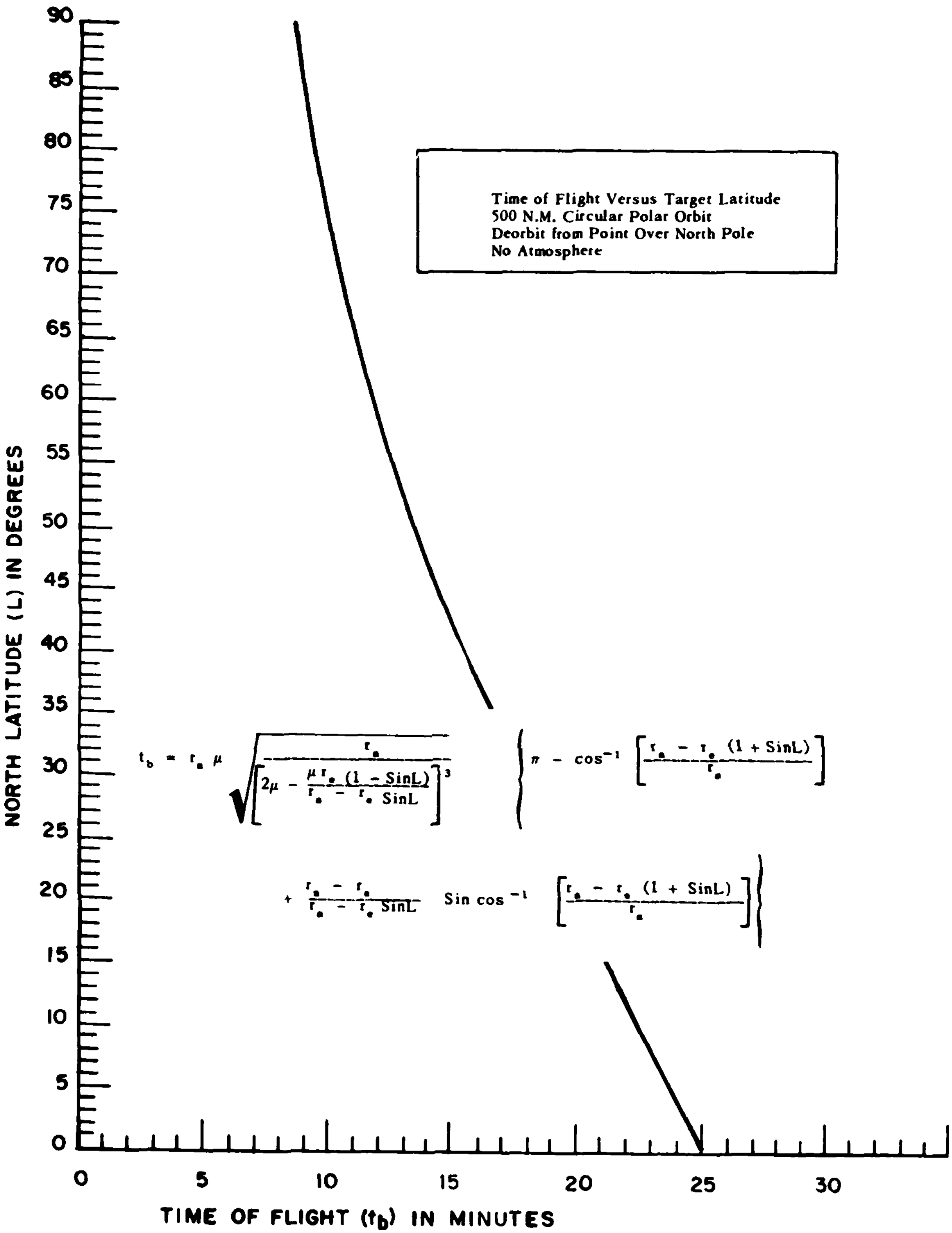


Figure A-6.



$$t_b = r_a \mu \sqrt{\frac{r_a}{\left[2\mu - \frac{\mu r_e (1 - \mathrm{Sin}L)}{r_a - r_e \mathrm{Sin}L}\right]^3}} \left\{ \pi - \cos^{-1}\left[\frac{r_a - r_e (1 + \mathrm{Sin}L)}{r_a}\right] + \frac{r_a - r_e}{r_a - r_e \mathrm{Sin}L} \mathrm{Sin}\cos^{-1}\left[\frac{r_a - r_e (1 + \mathrm{Sin}L)}{r_a}\right]\right\}$$

Figure A-7.

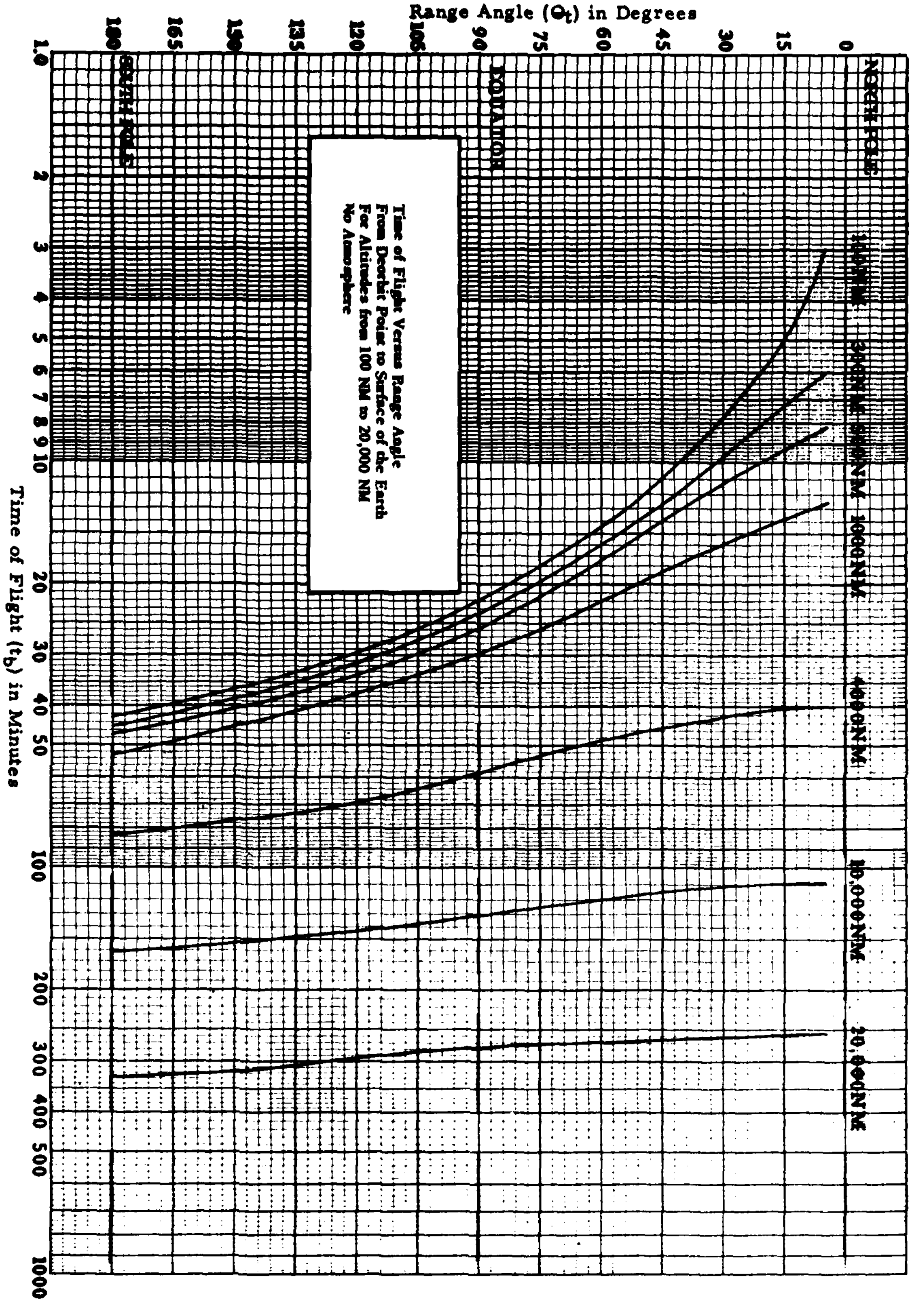

Range Angle (Θ_t) in Degrees
180 165 150 135 120 105 90 75 60 45 30 15 0
NORTH POLE
EQUATOR
SOUTH POLE
Time of Flight Versus Range Angle
From Deorbit Point to Surface of the Earth
For Altitudes from 100 NM to 20,000 NM
No Atmosphere
100 NM
300 NM
500 NM
1000 NM
4000 NM
10,000 NM
20,000 NM
1.0 2 3 4 5 6 7 8 9 10 20 30 40 50 100 200 300 400 500 1000
Time of Flight (t_b) in Minutes

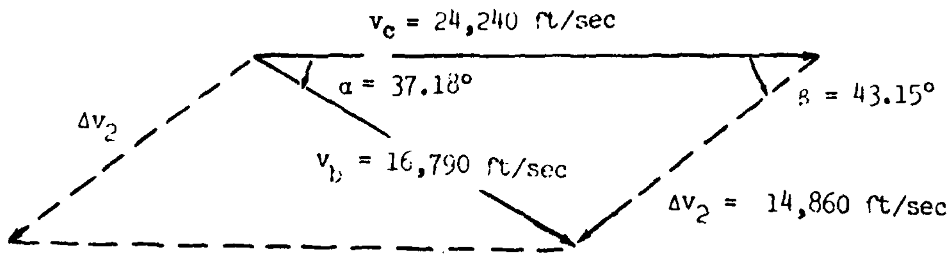
Figure A-8.

$$(360°/24\ \text{hrs})\frac{\text{hr}}{60\ \text{min}} = .25\ \text{deg/min}$$

$$\theta_r = (11.27\ \text{min})(.25\ \text{deg/min}) = 2.82°$$

$$\alpha = 40° - 2.82° = 37.18°$$

Now apply the Law of Cosines and determine the Δv_2 that must be applied to v_c in order to impact the target.



$$\Delta v_2 = (v_b^2 + v_c^2\ 2v_b\, v_c \cos\alpha)^{1/2}$$

$$\Delta v_2 = [(16{,}790\ \text{ft/sec})^2 + (24{,}240\ \text{ft/sec})^2 - 2(16{,}790\ \text{ft/sec} \times 24{,}240\ \text{ft/sec}\cos 37.18°)]^{1/2}$$

$$\Delta v_2 = [2.82 \times 10^8 + 5.88 \times 10^8 - 6.49 \times 10^8]^{1/2}$$

$$\Delta v_2 = [2.21 \times 10^8]^{1/2} = 14{,}860\ \text{ft/sec}$$

We must know the value of angle β so that we may determine the proper direction of Δv_2. The law of sines is preferred for this calculation, although we can use the law of cosines.

$$\sin\beta = \frac{v_b \sin\alpha}{\Delta v_2}$$

$$\sin\beta = \frac{16{,}790\ \text{ft/sec}\sin 37.18°}{14{,}860\ \text{ft/sec}}$$

$$\sin\beta = .6827$$

$$\beta = 43.15°$$

From the solution of this simplified object-from-object problem, it is apparent that such a calculation is not really simple. Other mathematical approaches and other operational problems are even more difficult. However, we can extend the theory presented to more difficult cases.

FUEL REQUIREMENTS

It is interesting to determine the propellant necessary to perform this maneuver. Assuming an I_{sp} = 450 seconds (a reasonable figure in the near future) and an initial weight of 10,000 pounds, compute the amount of propellant required, w_p.

$$\ln\left(\frac{W_1}{W_2}\right) = \frac{\Delta v}{I_{sp}g} = \frac{14{,}860\ \text{ft/sec}}{(450\ \text{sec})\ (32.2\ \text{ft/sec}^2)} = 1.025$$

$$\frac{W_1}{W_2} = 2.79$$

$$W_2 = \frac{10{,}000\ \text{lb}}{2.79} = 3{,}590\ \text{lb}$$

$$W_p = W_1 - W_2 = 6{,}410\ \text{lb}$$

This weight of propellant represents $\frac{W_p}{W_1} = \frac{6,400}{10,000}$ = 64.1 percent of the weight in orbit prior to maneuvering. By restricting the plane change (side range) to α = 7.18°, Δv_2 = 7,875 ft/sec, W_p = 4,190 lb, and $\frac{W_p}{W_1}$ = 41.9 percent.

In summary, recall that we gave the altitude of a satellite in a circular orbit and a time when the satellite was over the North Pole. A target was selected and no earth atmosphere was assumed. From the theory, we calculated the required velocity for a deorbiting object and predicted the position of the target at the time of impact. With this information and the use of the laws of sines and cosines, we calculated the magnitude of retrovelocity and the direction to deorbit on target. We computed the amount of propellant as well.

Appendix B

DETERMINATION OF THE ANGLE BETWEEN TWO ORBITAL PLANES

For space rendezvous to occur, both vehicles must simultaneously be in the same orbital plane and be at the same location in identical orbits.* For this discussion, we will assume that the orbital requirement has been satisfied independently of the orbital plane requirement. The maneuvering operations necessary to satisfy both requirements may occur in any sequence. However, we will discuss only the orbital plane requirement here.

In space rendezvous operations, we encounter the problem of reaching a specified orbital plane from either another orbital plane or a specific launch site. Determination of the plane change angle (α) required to change from one orbital plane to another, at the intersection of two planes, is complicated only by the requirement to use spherical geometry instead of the plane trigonometry to which we are more accustomed.

To simplify the discussion of the solution, we define the orbital plane of the vehicle as the initial plane, and we define the desired new orbital plane as the final plane. Since any launch must enter, initially, an orbital plane that is dependent on the launch site latitude (L), launch azimuth, (β), and time and date of launch, the above definitions will suffice.

For the case of the initial plane and a final plane, the solution is dependent on knowing or estimating the right ascension (Ω) and inclination (i) angles of the two planes. For the case of a final plane and a launch site, the solution is dependent on the Ω and i angles for the plane, time between launch and passage of the orbital plane over the launch site, and the launch latitude and azimuth.

Two examples given illustrate the mathematics involved and the approximation techniques available for obtaining solutions. In the first example, we will show that only the difference in right ascension angles ($\Delta\Omega$) is required. The specific value of Ω need not be known.

CASE I

We want to maneuver a vehicle in orbit 1 into the plane of orbit 2. To accomplish this maneuver, we must know the value of the plane change angle.

Orbital Parameters

	Orbit 1	Orbit 2
Ω	30°	45°
i	28°	30°

From spherical trigonometry, we can determine the plane change angle from:

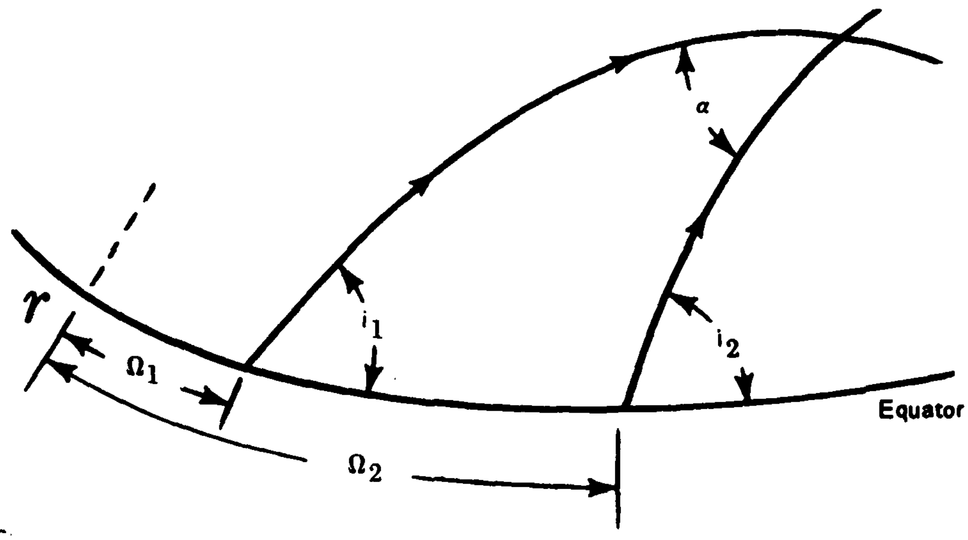
$$\cos\left(\frac{a}{2}\right) = \frac{\cos\dfrac{i_2 - i_1}{2}}{\cos(x)} \cos\frac{\Omega_2 - \Omega_1}{2}$$

*Identical orbits occur when the magnitude and orientation of the major axes (2a) are the same and the eccentricities (e) are equal

Where the angle x, a dummy angle used in the calculation, is determined from:

$$\tan (x) = \frac{\cos \frac{i_2 + i_1}{2}}{\cos \frac{i_2 - i_1}{2}} \tan \frac{\Omega_2 - \Omega_1}{2}$$

The geometry of the problem is as shown below.



From the equations and the geometry it is apparent that: only the difference ($\Delta\Omega$) in Ω is required; it will always be positive; and it can be reduced to an angle between 0 and 180 degrees. Therefore, for this problem $\Delta\Omega$ is 15°.

Substituting to determine the angle x.

$$\tan x = \frac{\cos \frac{30° + 28°}{2}}{\cos \frac{30 - 28}{2}} \tan \frac{15°}{2} = \frac{\cos (29°)}{\cos (1°)} \tan (7.5°)$$

$$= \frac{0.8746197}{0.9998477} \times (0.131652) = 0.1151629$$

Therefore, $x = 6° \ 34'$ and $\cos x = 0.9934727$

Substituting to determine the plane change angle α:

$$\cos (\frac{a}{2}) = \frac{\cos \frac{30° - 28°}{2}}{0.9934727} \cos (7.5 °)$$

$$= \frac{0.9998477}{0.9934727} + (0.9914449) = 0.997807$$

Therefore, $\frac{\alpha}{2} = 3° \ 48'$ or $\alpha = 7° \ 36'$.

The required plane change angle is 7° 36′.

CASE II

A spacecraft on a launch vehicle at Cape Kennedy (L = 28°) must attain orbit in a plane that has a southwest-northeast orientation and an inclination angle of 30°. Ideally, the vehicle would be launched and injected along the required azimuth angle, when the desired orbital plane was directly over the launch site. However, this is a very difficult timing problem. The spacecraft/launch vehicle combination has limited velocity change capabilities, which narrows the launch window. A launch occurring within this window of time will permit injection into an initial parking orbit and a subsequent plane change maneuver to the desired orbital plane.

Because of launch delays, the launch site passed through the desired final orbital plane eight minutes prior to lift-off. Since the launch vehicle guidance was programmed prior to launch, the spacecraft will be injected into an initial parking orbit, and then it must make a plane change to the desired final orbital plane. The programmed launch azimuth angle is 79°. What is the value of the plane change angle required to complete the mission? The apparent unknowns are the initial orbit inclination angle, and the $\Delta\Omega$.

To find the inclination angle of this parking orbit, use the equation for inclination angle of the launch:

$$\cos i = \cos (L) \sin (\beta) \approx \cos (28°) \sin (79°) = 0.8660$$

Therefore, $i = 30°$ for the parking orbit. It is possible to approximate $\Delta\Omega$ using the earth's rate of rotation.

$$\text{rotation rate} = \frac{360°}{24 \text{ hrs}} = \frac{360°}{(24 \times 60) \text{ min}} = 0.25°/\text{min}$$

For the situation of equal inclination angles, the $\Delta\Omega$ equals the time between launch and passage of the final orbital plane overhead, multiplied by the earth's rotation rate:*

$$\Delta\Omega = 0.25°/\text{min} \times 8 \text{ min} = 2°$$

By making use of the same equations used for case one, the plane change angle α will be approximately 1°.

LIST OF SYMBOLS

α — Plane Change Angle
β — Azimuth Angle
$\Delta\Omega$ — Difference in Right Ascension Angles
Ω — Right Ascension Angle
i — Inclination Angle
L — Angle of Latitude
x — Dummy Angle

Subscripts

1. Left hand orbital plane.
2. Right hand orbital plane.

Superscripts

* Dummy Variable

---

*For the case of nonequal inclination angles, it is necessary to add a term $\Delta\Omega^*$.

$$\Delta\Omega^* = \Omega^*_1 \mp \Omega^*_2 \quad \begin{cases} - & \text{Both inclination angles in the same quadrant} \\ + & \text{Inclination angles in different quadrants} \end{cases}$$

$\Delta\Omega^*$ must be positive

To find the Ω^*'s use

$$\sin \Omega^*_1 = \tan (L) \cot (i_1)$$
$$\sin \Omega^*_2 = \tan (L) \cot (i_2)$$

Then the total $\Delta\Omega$ equals the $\Delta\Omega$ due to the earth's rotation plus the $\Delta\Omega^*$ due to the difference in inclination angles.

Appendix C

SIGNS AND SYMBOLS

MATHEMATICAL SIGNS AND SYMBOLS

$\pm$	plus or minus, positive or negative
$\neq$	is not equal to
$\equiv$	is identical to
$\approx$	approximately equal to
$>$	greater than
$\geqq$	greater than or equal to
$<$	less than
$\leqq$	less than or equal to
$\backsim$	similar to
$\propto$	varies as, proportional to
$\rightarrow$	approaches as a limit
∞	infinity
$\therefore$	therefore
$\sqrt{\ }$	square root
$\sqrt[n]{\ }$	nth root
a^n	nth power of "a"
a^{-n}	reciprocal of nth power of a $= (\frac{1}{a^n})$
log, $\log_{10}$	common logarithm
ln, $\log_e$	natural logarithm
n°	n degrees
n'	n minutes; n feet
n''	n seconds; n inches
f(x)	function of x
Δx	increment of x
dx	differential of x
∂x	partial differential of x
Σ	summation of
$\int$	symbol for integration

GREEK ALPHABET

Alpha	A α	Iota	I ι	Rho	P ρ
Beta	B β	Kappa	K κ	Sigma	Σ σ
Gamma	Γ γ	Lambda	Λ λ	Tau	T τ
Delta	Δ δ	Mu	M μ	Upsilon	Y υ
Epsilon	E ϵ	Nu	N ν	Phi	Φ ϕ
Zeta	Z ζ	Xi	Ξ ξ	Chi	X χ
Eta	H η	Omicron	O o	Psi	Ψ ψ
Theta	Θ θ	Pi	Π π	Omega	Ω ω

GLOSSARY OF SYMBOLS

a	semimajor axis of ellipse; average linear acceleration
b	semiminor axis of ellipse
bo	subscript for burnout conditions
c	distance between focus and center of ellipse
e	base of natural logarithm, = 2.718
g	<u>local</u> acceleration due to gravity, = 32.2 ft/sec^2 at the surface of the earth
h	altitude, height above surface of earth
h_a	altitude of apogee
h_p	altitude of perigee
i	angle of inclination
p	electric power
r	radius length; mixture ratio
r_a	radius to apogee
r_e	radius of earth, = 3440 NM = 20.9×10^6 ft
r_p	radius to perigee
s	linear displacement
t	time in seconds
u	eccentric anomaly
v	linear speed, velocity magnitude
v_e	nozzle exit velocity
$\triangle v$	increment or change of speed
w	work in foot-pound force
A	area
A_e	nozzle exit area
C_D	coefficient of drag
C_l	coefficient of lift
D	atmospheric drag
E	specific mechanical energy in ft^2/sec^2

F	focus of ellipse; force in pounds; thrust in pounds
G	Universal Gravitational Constant, = 10.69×10^{-10} ft^3/lb mass-sec^2
H	specific angular momentum in ft^2/sec
I_{sp}	specific impulse in seconds
I_t	total impulse in pound-seconds
KE	kinetic energy
L	aerodynamic lift; latitude
M	mass in slugs
$\dot{M}$	mass flow rate
NM	nautical miles, = 6080 feet
P	period of revolution; pressure
PE	potential energy
Q	ballistic trajectory parameter; reactor thermal power
W	weight in pounds force
$\dot{W}$	weight flow rate
α	average angular acceleration
δ	very small change or error
ϵ	eccentricity; expansion ratio
η	electrical efficiency
θ	angular displacement
λ	longitude
μ	gravitational parameter, = 14.08×10^{15} ft^3/sec^2 for earth
ν	true anomaly; heat transfer efficiency
π	conversion constant, = 3.1416; π radians = 180°
ρ	atmospheric density, slugs/ft^3
ϕ	flight path angle, elevation angle of velocity vector
ψ	free flight range angle
ω	argument of perigee; average angular speed
Δ	increment of
ψ	thrust to weight ratio

1986-631-028/20652 AUGAFS,AL(860747)3000

END
DTIC
FILMED
4-86